THE

PROCEEDINGS

OF

THE PHYSICAL SOCIETY

FROM JANUARY 1946 TO NOVEMBER 1946

VOLUME 58
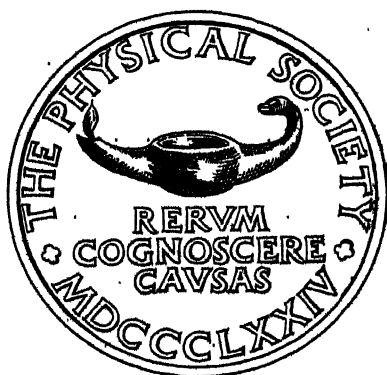
OFFICERS AND COUNCIL, 1946-47

PRESIDENT

D. BRUNT, M.A., Sc.D., F.R.S.

VICE-PRESIDENTS
who have filled the Office of President

C. H. LEES, D.Sc., F.R.S. (1918–20)
Sir FRANK SMITH, G.C.B., G.B.E., D.Sc., LL.D., F.R.S. (1924–26)
Sir OWEN RICHARDSON, M.A., D.Sc., F.R.S. (1926–28)
W. H. ECCLES, D.Sc., M.I.E.E., F.R.S. (1928–30)
A. O. RANKINE, O.B.E., D.Sc., F.R.S. (1932–34)
The Right Hon. LORD RAYLEIGH, M.A., Sc.D., L.L.D., F.R.S. (1934–36)
T. SMITH, M.A., F.R.S. (1936–38)
ALLAN FERGUSON, M.A., D.Sc. (1938–41)
Sir CHARLES DARWIN, K.B.E., M.C., M.A., Sc.D., F.R.S. (1941–43)
E. N. da C. ANDRADE, Ph.D., D.Sc., F.R.S. (1943–45)

VICE-PRESIDENTS

N. F. MOTT, M.A., F.R.S.
Sir EDWARD APPLETON, G.C.B.,
  M.A., D.Sc., LL.D., F.R.S.

A. I. PHILPOT, C.B.E., M.A., B.Sc.
H. R. ROBINSON, D.Sc., Ph.D., F.R.S.

HONORARY SECRETARIES

W. JEVONS, D.Sc., Ph.D. (Business)   J. H. AWBERY, M.A., B.Sc. (Papers)

HONORARY FOREIGN SECRETARY

E. N. da C. ANDRADE, Ph.D., D.Sc., F.R.S.

HONORARY TREASURER

H. SHAW, D.Sc.

HONORARY LIBRARIAN

L. C. MARTIN, D.Sc.

ORDINARY MEMBERS OF COUNCIL

E. R. DAVIES, O.B.E., B.Sc.
W. B. MANN, Ph.D.
W. D. WRIGHT, D.Sc.
B. CHALMERS, D.Sc., Ph.D.
W. S. STILES, D.Sc., Ph.D.
C. H. COLLIE, M.A., D.Phil.

J. D. BERNAL, M.A., F.R.S.
G. I. FINCH, M.B.E., D.Sc., F.R.S.
D. ROAF, M.A., D.Phil.
A. C. G. MENZIES, M.A., D.Sc.
R. E. PEIFRLS, C.B.E., M.A., D.Sc., F.R.S
F.C. TOY, D.Sc.

COLOUR GROUP

Chairman
R. K. SCHOFIELD, M.A., Ph.D.

Honorary Secretary
W. D. WRIGHT, D.Sc.

OPTICAL GROUP

Chairman
Instr.-Capt. T. Y. BAKER, B.A., R.N. (ret.).

Honorary Secretary
E. W. H. SELWYN, B.Sc.

LOW-TEMPERATURE GROUP

Chairman
Sir ALFRED EGERTON, M.A., Sec. R.S.

Honorary Secretary
G. G. HASELDEN, B.Sc.

CONTENTS

Contents

PROCEEDINGS AT THE MEETINGS
OF THE PHYSICAL SOCIETY

SESSION 1945-46

19 September 1945

The twenty-third meeting of THE COLOUR GROUP, at Imperial College, London S.W.7. Dr. R. K. Schofield was in the Chair.

A lecture, with colour films and demonstrations, on " Methods for investigating colour discrimination in insects " was delivered by Dora R. Ilse, D.Phil.

---

28 September 1945

The eighteenth meeting of THE OPTICAL GROUP, at Imperial College, London S.W. 7. Professor L. C. Martin was in the Chair.

A lecture on " Evolution of the dividing engine " was delivered by E. Wilfred Taylor. A film entitled " Motion-picture records of the whole sky " was shown.

---

5 October 1945

Science Meeting, at Imperial College, London S.W. 7. The President, Professor D. Brunt, was in the Chair.

David William Elson (transferred from Student Membership) and Allen Goodrich Shenstone were elected to Fellowship.

It was announced that the Council had elected the following to Student Membership : Robert Batchelor, Michael Julian Maurice Bernal, William Ralph Blow, James William Crawford, Albert Julius Elleman, D. J. Hudson, Susanna Korn, Charles W. Morley, Michael William Ovenden, John Oulton Quantrell, William Bernard Richardson, Thomas Templeton, Hyman Woolfe.

The following papers were read and discussed :

" A method for obtaining small mechanical vibrations of known amplitude ", by D. H. Smith, D.Sc.

" The photographic action of x rays ", by S. R. Pelc, Ph.D.

" The hysteresis of the angle of contact of mercury ", by the Rev. G. D. Yarnold, M.A., D.Phil.

---

17 October 1945

Inaugural meeting of THE LOW-TEMPERATURE GROUP, at The Royal Society, Burlington House, London W. 1. Professor Sir Alfred Egerton was in the Chair.

A draft constitution of the Group was discussed and adopted. The Chairman, Vice-Chairman, Secretary and Committee of the Group for 1945–46 were elected.

An address on " Some technical problems of gas separation " was given by A. M. Clark.

---

19 October 1945

Science Meeting, at The Royal Society, Burlington House, London W.1. The President, Professor D. Brunt, was in the Chair.

The following were elected to Fellowship : John Boston Beecher, Raymond Martin Boorman, Francis Haddon Bowen, Arthur Breeze-Carr, Pierre Albert Bricout, Henry Arthur Druett, Eric Paul George, Robert Edward William Iggkeden, Henry Lipson,

Alfred Charles Bernard Lovell, Harold Overton, Ernest Ower, Owen George Weller, Robert Lewis Woolley and Walter Zehden.

It was announced that the Council had elected the following to Student Membership : Freda Betty Freeman, Audrey Marian Glauert, Eileen Beryl Harriss, William Maurice, Gordon von Noehlsen, Philip Leighton Parsons, Derrick Arnold Patient, James Clifford Rees, Keith Robinson, Geoffrey Donald Sims, Hilary Tadman, John William George Taylor, David John Cabena Yates.

A lecture on " The spectral structure of the inert gases " was delivered by Professor Bengt Edlén.

---

14 November 1945

The twenty-fourth meeting of THE COLOUR GROUP, at the Lighting Service Bureau, Savoy Hill, London W.C. 2.   Dr. R. K. Schofield was in the Chair.

A lecture on " Colorimetry in the glass industry " was given by J. G. Holmes, B.Sc., and was followed by an informal discussion.

---

6 December 1945

Science Meeting, at the Science Museum, London S.W. 7.   The President, Professor D. Brunt was in the Chair.

The following were elected to Fellowship : Bruce Henry Burgess, Norman Edward Cusack, Leslie Bernard Hunt, Charles Lessner, John Millar Meek, Joseph Tunstead.

It was announced that the Council had elected the following to Student Membership : Jerzy Adam, Edward Walter Bastin, Norman Frank Blight, Patrick Bomyer, Raymond Bowers, Robert James Helton Branthwaite, Laurence Reade Cooper, Dennis Peter Craft, Raymond Charles Curnow, William Kenneth Donaldson, Jósef Dlugosz, John Lewis Fanner, Doris Evelyn Gould, Peter William Hoare, Robert Tudor Jarman, Frederick Alexander Jenner, Eric Walter Lee, Peter Donald Lomer, Frederick John Brian Oliver, Donald William Pashley, John Rolfe, Catherine Patricia Strang, Derek Alan Webber, William Leslie Wilcock.

The Charles Chree Medal and Prize were presented to John A. Fleming, Sc.D., who afterwards delivered the third Charles Chree Address, taking as his subject "Geomagnetic secular variations and surveys ".

---

12 December 1945

Science Meeting, at the Science Museum, London S.W. 7.   The President, Professor D. Brunt, was in the Chair.

The following were elected to Fellowship : Madge G. Adam, John Alois Becker, Alan West Brewer, Geoffrey Fernis Dixon, John Macfie Holm, Gwyn Owain Jones, Edward K. Kaprelian, Alec Herbert King, Samuel Thomas Lunt, Denis Moody, Reginald Charles Moore, Neil Pentland, Alan Turner Pickles, John Turton Randall, Arthur Rushton, John Moffett Cuthbert Scott, Sudarshan Prasad Sinha, David Gurney Arnold Thomas, Alfred René John Paul Ubbelohde, Morgan Thomas Watkins, Bernard Nelson Watts.

The twenty-second (1945) Duddell Medal was presented to Professor J. T. Randall, D.Sc., F.R.S., who afterwards gave a lecture, with demonstrations, on his work on phosphors and the magnetron.

The first (1945) Charles Vernon Boys Prize was presented to A. H. S. Holbourn, M.A., D.Phil., who afterwards gave an account of his work on " The measurement of the angular momentum of radiation ".

---

14 December 1945

The nineteenth meeting of THE OPTICAL GROUP, at the Science Museum, London S.W. 7 Professor L. C. Martin was in the Chair.

An informal discussion on " The coating of optical components to reduce reflection " was opened by K. M. Greenland, Ph.D.

---

19 *December* 1945

The twenty-fifth meeting of THE COLOUR GROUP, at Imperial College, London ·S.W. 7.
Dr. R. K. Schofield was in the Chair.
   A lecture on " Colour discrimination, the visibility curve and the trichromatic theory "
was given by W. S. Stiles, D.Sc., Ph.D., and was followed by an informal discussion.

---

16 *January* 1946

The second meeting of THE LOW-TEMPERATURE GROUP, at the Science Museum, London
S.W. 7.   Sir Alfred Egerton was in the Chair.
   An informal discussion on " The needs of industry for investigations to provide
fundamental data " was opened by P. M. Schuftan.

---

30 *January* 1946

The twenty-sixth meeting of THE COLOUR GROUP, at Imperial College, London S.W. 7.
Dr. R. K. Schofield was in the Chair.
   The following papers were read and discussed :
" A colorimeter with six matching colours ", by R. Donaldson, M.A.
" A direct-reading photoelectric spectrophotometer ", by T. B. Davenport, B.Sc.

---

8 *February* 1946

Science Meeting, at the Science Museum, London S.W. 7.   The President, Professor D.
Brunt, was in the Chair.
   William Ernest Bennett, James William Jeffery and Peter Longdale Temple were
elected to Fellowship.
   It was announced that the Council had elected the following to Student Membership
Richard Sidney Babbs, Peter Frederick Barker, R. J. Fletcher, Gordon David Reid Granick,
George William Hamstead, Richard Maxwell Harkness, Keith William Ogilvie, Saxon
Maclean Poole, Ruth Anne Lund Roberts, Stanley Keith Runcorn, Sydney Eugene
Veronique.
   The second Rutherford Memorial Lecture was delivered by Professor J. D. Cockcroft,
C.B.E., Ph.D., F.R.S., who took as his subject " Rutherford : Life and work after the year
1919, with personal reminiscences of the Cambridge period ".

---

15 *February* 1946

Science Meeting, at the Science Museum, London S.W. 7. · The President, Professor D.
Brunt, was in the Chair.
   The following were elected to Fellowship, the last eighteen being transferred from
Student Membership : Brebis Bleaney, Cecil Leonard Boltz, Kenneth George Budden,
Bohdan Medusa Cwilong, Maurice Charles Pierre Desirant, David Henry Follett, John
Samuel Forrest, Harold Arthur Freeman, William Anthony Greenwood, Geoffrey Thomas
Harris, Eric William Masson Heddle, William Campbell Heselwood, John Riley Holt,
John Anthony Wynne Huggill, Lamek Hulthén, Tudor Stanley George Jones, Walter
Alfred Langmead, Constance Marie Lovett, Malcolm McCaig, Ladislaus Marton, Frank
Rodney Moody, Ivan Brian Perrott, Maurice Henry Lecorney Pryce, Edward Rollinson,
Joseph Rotblat, John Arthur Saxton, Helen Scouloudi, Albert Robert Service, Edward
Samuel Shire, G. H. St. Claire-Johnson, Harold Henry Watson, Robert Foster Wilkinson,
Robert Charles Gooding Williams, George Andrew Park Wyllie ; Harry Edward Airey,
Peter Milton Barham, Francis Julian Bradshaw, Basil Hugh Briggs, Peter Crombie Calder,
Miriam Gilbert, Geoffrey Blakeley Greenough, Douglas Hardy, James Morden Hough,

Betty Hunt, Robert Henry Kay, Joan Kelly, Dennis Pallant, Percy Charles Ruggles, Alexander Rawson Stokes, Ronald Tiffen, Theodore James Tulley, Anthony Horace Willbourn.

It was announced that the Council had elected the following to Student Membership : John Albert Bleach, Derek Coulthard, Raymond Charles Curnow, Jan Marcin Engel, Basil Arthur Fancourt, Tom Fawzi, Alan Neville Gent, Margaret Enid Hall, Arthur Emlyn Dawson Harris, Paul Francis Hatherly, Kenneth Roy Heath, Thomas Jermyn, Thomas Reginald Maynard, Arthur Cyril Moore, Bethune William Neill, Lewis Percival Newman, Reginald Percy Penrose, Eric Pitts, Thomas Noel Reynolds, Raymond Jeffrey Slaughter, Dorothy Jean Smith, Richard Thomas Albert Standford, William John Symes, Dewi Gwynfa Bailey Thomas, Krishnarao Govindarao Torgal, Patrick·Joseph Treanor, Kashi Prasad Verma, Margaret Elizabeth Weiss, Peter Charles Williamson, Frederick Ronald Young.

A lecture on " The physical and rheological properties of glass " was delivered by G. O. Jones, M.A., Ph.D., and was followed by an informal discussion.

---

1 *March* 1946

The twentieth meeting of THE OPTICAL GROUP, at the Science Museum, London S.W. 7. Instr.-Capt. T. Y. Baker, was in the Chair.

The following papers were read and discussed:

" The angle-control aspherizing machine ", by C. R. Burch, F.R.S.

" An improved type of Schmidt camera ", by E. H. Linfoot, B.A., D.Phil.

" Decentred aspheric plates " by E. H. Linfoot, B.A., D.Phil.

---

13 *March* 1946

The twenty-seventh meeting of THE COLOUR GROUP, at the Science Museum, London S.W. 7. Dr. R. K. Schofield was in the Chair.

A lecture on " Light absorption and elementary wave mechanics " was delivered by E. J. Bowen, M.A., F.R.S., and was followed by an informal discussion.

The meeting was preceded by the sixth Annual General Meeting of the Colour Group, for the presentation of the Committee's report on the work of the Group in 1945–46 and for the election of Officers and Committee for 1946–47.

---

13 *March* 1946

The third meeting of THE LOW-TEMPERATURE GROUP, at the Works of the British Oxygen Company, Wembley, Middlesex.

The older and the newer processes for preparing liquid and gaseous oxygen were described and inspected.

---

15 *March* 1946

Science Meeting, at Imperial College, London S.W. 7. The President, Professor D. Brunt, was in the Chair.

The following were elected to Fellowship, the last three being transferred from Student Membership : William Ernest Doran, Eric Basil James, Robert Latham, Alan Lawson Little, Maurice Henri Léonard Pirenne, Ralph Ellis White ; John Gadsby, Alfred Peter Henry Jennings, Ronald William Shephard.

It was announced that the Council had elected the following to Student Membership : Desmond Martin Slingsby Bagguley, Ian David Leonard Ball, John James Knight, Gordon Kingsley Monks, Raymond Paul Rose, Leyland Whylock Shawe, Peter Sleightholm, Robert W. Taylor, William Harris Thorning, Raymond Marriage Wallis, Gwyn Clement Williams, Geoffrey Leonard Wilson.

The following papers wcre read and discussed:

" The relation between the brightness and temperature of a total radiator ", " The ratio of the new international lumen to the light-watt ", " The determination of the initial temperature of a cooling total radiator from measurements of the spectral distribution of the energy emitted during cooling ", and " Light emission during cooling of a Planckian radiator ", by E. F. Caldin, M.A., B.Sc.

" Band systems in the spectrum of nitrogen ", by A. G. Gaydon, D.Sc., and Mme R. Herman, D. ès Sc.

" The construction and use of a ' fly's eye ' for assisting x-ray structure analysis ", by A. R. Stokes, M.A., Ph.D.

---

29 *March* 1946

The twenty-first meeting of THE OPTICAL GROUP, at Imperial College, London S.W. 7 Instr.-Capt. T. Y. Baker was in the Chair.

The following demonstrations were given:

" Effects of ultra-violet radiations of different wave-length upon luminescence ", by J. A. Kitchener, Ph.D., and C. G. A. Hill, B.Sc.

" Comparator for checking curvature of unpolished lenses ", by R. S. Clay, D.Sc.

" Measurement of small linear motions by optical methods ", by R. G. W. Hunt, B.Sc.

A film of " Meiosis of sperm cells of grasshopper ", taken with a phase-contrast microscope, was shown by E. H. Linfoot, B.A., D.Phil.

---

8 *April* 1946

Science Meeting, held jointly with the Royal Meteorological Society, at the Royal Institution, London W. 1. The Chair was taken by Mr. G. Manley, M.A. (President, R. Met. Soc.) during the opening session, and afterwards by Professor D. Brunt (President, Phys. Soc.).

The following were elected to Fellowship, the last five being transferred from Student Membership : Eric Charles Baxter, John Douglas Cockcroft, Lionel J. Collier, John Hudson Davies, Alexander Constantine Denisoff, Percy John Edwards, Norman Feather, Colin Charles Hall, John Henry Jeffree, Richard Harday Knight, Edward Arthur Lovell, John Blackburn Murgatroyd, Samuel Tolansky, Reinhold Adolf Philipp Wertheim ; Roy Drakeley Bruce, Geoffrey George Parfitt, Trevor Albert John Stocker, Donald Williams, Geoffrey Philip Lewis Williams.

A Conference on " Meteorological factors in radio-wave propagation " was opened by Sir Edward Appleton, K.C.B., Sc.D., F.R.S.

---

26 *April* 1946

The twenty-second meeting of THE OPTICAL GROUP, at Imperial College, London S.W. 7. Instr.-Capt. T. Y. Baker was in the Chair.

The following papers were read and discussed:

" Schwarzschild-Ritchey-Chrétien reflecting microscopes ", by C. R. Burch, F.R.S.

" A Lenouvel interferometer " (with demonstration), by F. T. Bannister, W. J. Bates and C. R. Burch, F.R.S.

The meeting was preceded by the fifth *Annual General Meeting of the Optical Group* for the presentation of the Committee's report on the work of the Group in 1945–46 and for the election of Officers and Committee for 1946–47.

16 *May* 1946

Extraordinary General Meeting and *Annual General Meeting*, at the Royal Institution,. London W. 1. The President, Professor·D. Brunt, was in the Chair.

It was resolved " That Article 36 of the Articles of Association be suspended for a period of one year ".

The minutes of the previous Annual General Meeting, held on 23 May 1945, were read and confirmed.

The reports of the Council and the Honorary Treasurer and the Annual Accounts for 1945 were adopted.

The Officers and Council and the Auditors for 1946–47 were elected.

Votes of thanks were accorded to the Rector and Governing Body of Imperial College, the President and Council of the Royal Society, the Managers of the Royal Institution, and the Director of the Science Museum for excellent accommodation at meetings ; to the Royal Commission for the Exhibition of 1851 and Dr. Evelyn Shaw for the office and library accommodation at 1 Lowther Gardens, Prince Consort Road, London S.W. 7 ; to Dr. A. F. C. Pollard for preparing the U.D.C. Index Slips for the *Proceedings* ; and to the retiring Officers and Council.

---

16 *May* 1946

Science Meeting, at the Royal Institution, London W. 1. The President, Professor D. Brunt, was in the Chair.

The following were elected to Fellowship, the last four being transferred from Student Membership : Edgar Harold Walter Banner, William Stanley Bowler, John Espenett Caffyn, Maxwell Dauer, Dai Ashton Davies, Norman David Imrie, John Arnold Lane, Sidney Barron Osborn, William Henderson Ramsey, Walter L. Stern, Edward Alfred Stewardson, Norman Thorley, Richard Franz Jules Vogel, William John Whitehouse ; Arthur Norman Hunter, Richard Claude Lane, Donald Vernon Osborne, Nigel Graham Trott.

It was announced that the Council had elected the following to Student Membership : Ian Alexander Gatenby, Lewis Roy Griffin, John Williamson MacAnuff, Audrey Doris Stuckes, David George Ware.

The first (1946) Holweck Prize was presented to Professor C. Sadron, who afterwards ·delivered the Holweck Discourse on " Some physical properties of long-chain molecules ".

His Excellency the French Ambassador (M. René Massigli) and distinguished repre sentatives of the Société Française de Physique were present.

---

23 *May* 1946

The twenty-eighth meeting of THE COLOUR GROUP, at the Royal Photographic Society, London S.W. 7. Dr. R. K. Schofield was in the Chair.

A paper on " The colour sensitivity of the retina within the central fovea of man " was read by L. C. Thomson, M.D., and W. D. Wright, D.Sc., and was followed by an informal discussion.

---

28 *June* 1946

The twenty-third meeting of THE OPTICAL GROUP, at the Northampton Polytechnic, London E.C. 1. Instr.-Capt. T. Y. Baker was in the Chair.

The following papers were read :

" A survey of pupil diameters at background brightnesses ranging from darkness to full daylight " and " The night focus of the eyes ", by J. Tunstead, Ph.D.

" Acrylic artificial eyes ", by A. Kingston and E. F. Fincham.

" Some notes on visual acuity ", by Professor L. C. Martin, D.Sc.

" The variation of visual acuity with pupil diameter ", by E. W. H. Selwyn, B.Sc.

<center>5 *July* 1946</center>

Science Meeting, at the Science Museum, London S.W. 7.   Dr. W. S. Stiles was in the Chair.

The following were elected to Fellowship, the last three being transferred from Student Membership :   Jack Kenneth Barraclough, Hugh Cecil Binstead, William Bleloch, Alan Hugh Boud, Eric Harold Boyd, Denis Frederick Bracher, Herbert Walter Davies, William Cusack Fahie, John Wilson Findlay, Ramon Enrique Gaviola, Merryn Geoffrey Harwood, Stephen Patric Francis Humphreys-Owen, Egil A. Hylleraas, Mohamed Ibrahim Malik, Joseph A. Manché, Vikram A. Sarabhai, George Boris Townsend, Donald Whittaker, Alexander Wood ;   Herbert Cairns Bolton, Audrey May Brasnett Douglas (*née* Parker), Ronalde Philip Le Blanc.

It was announced that the Council had elected the following to Student Membership : Norman Herbert Edgar Ahlers, Donald Kenneth Ashpole, Norman Walter Bellwood, Kenneth Francis Bishop, Norman Charles Featherstone Chappell, Geoffrey Wallwork Eastwood, Gunther Eichholz, Michael Leslie Ferrar, Cyril Owen Green, William Ronald Loosemore, Horace Manley, William Mallinson McIsaac, Richard William Stevens.

The following demonstrations were given :

" A simple optical model demonstrating the principle of the Bragg x-ray spectrometer " and " A mechanical model illustrating the uranium chain reaction ", by F. A. B. Ward, B.Sc.

The following papers were read and discussed :

" The absorption of sound in a homogeneous porous medium ", " An apparatus for accurate measurement of the acoustic impedance of sound-absorbing materials ", and " The propagation of sound between walls of porous material ", by R. A. Scott, Ph.D.

" Cathode-luminescence :   Part I. Growth and decay processes ;   Part II. Current saturation and voltage effects ;   Part III. Discussion of results ", by J. W. Strange, Ph.D., and S. T. Henderson, M.A., Ph.D.

<center>22–27 *July* 1946</center>

At the Cavendish Laboratory, Cambridge.   *International Conference* on Fundamental Particles and Low Temperatures, opened by Professor Neils Bohr, For.Mem.R.S. (Copenhagen).

REPORT OF COUNCIL FOR THE YEAR
ENDED 31 DECEMBER 1945

INTRODUCTORY AND GENERAL

In many ways 1945 was a remarkable year in the history of the Society. The Holweck Prize was successfully established, the' Low-Temperature Group was inaugurated, and the heavy work of preparation for the resumption of the Exhibition of Scientific Instruments and Apparatus in January 1946 had to be undertaken in what were little better than war-time conditions. The membership again showed a satisfactory expansion, and the financial position was maintained with no additional cost to the members and subscribers.
. The continued generosity of the Royal Commission for the Exhibition of 1851 in providing the Society with excellent accommodation at 1 Lowther Gardens is gratefully acknowledged. In the summer it became necessary to take immediate steps to replace most of the office and library furniture, which was the property of the Royal Commission and the Institute of Physics, and the opportunity was taken of adding to the Library one large room which will eventually also be used for meetings of the Council and committees. The process of re-equipment will be slow and costly, but the result will be much to the advantage of all concerned, visitors and staff alike. In the first half of 1945 staff difficulties were even worse than in previous years, and it was not until the later months that new appointments and additions to the staff became possible.

The least satisfactory feature of the work to be reported is in respect of publications ; meagreness of the official allowance of paper and shortage of labour and equipment at the printers' works are the chief difficulties. By an exhausting effort a small edition of the Exhibition Catalogue was produced in time, but the long and increasing delays in the printing of the *Proceedings* and the *Reports on Progress in Physics* have caused deep concern. The Officers and Council are taking all steps available to them to bring about some acceleration.

HOLWECK MEDAL AND PRIZE

In January 1945' the President, Prof. E. N. da C. Andrade (later Foreign Secretary) issued to the Fellows and friends of the Society an appeal for contributions for the foundation of an annual prize in the interests of the closest co-operation with French physics and as a memorial to Dr. Fernand Holweck, the distinguished Director of Research at the Institut du Radium in Paris, who died at the hands of the Gestapo, and to other physicists who met their deaths or suffered great privation during the German occupation of France. A sum of nearly £900 was collected, and the Council decided that it should be so invested as to enable a Holweck Prize of £100 and a parchment certificate to be awarded annually for ten years to a French and to an English physicist alternately for distinguished work in experimental physics. The presentation to the French winner is to be made in London, beginning in 1946, and that to the British winner in Paris, beginning in 1947. The scheme evoked the liveliest interest and profoundest gratitude in French circles, and the Société Française de Physique, whose intimate co-operation in the selection of the prize-winners is to be sought, has founded a Holweck Medal, to be presented to each recipient of the Holweck Prize.

MEETINGS

During the year eleven Science Meetings took place, in addition to those of the Groups, and the average attendance was probably larger than in any previous year for which comparable data are available.

At three of the meetings, at Imperial College, papers and demonstrations were presented and discussed. Three meetings were held at the Science Museum, one of them for a discussion on the electron microscope and its uses (*Proceedings*, 57 (1945), p. 564), and two for special addresses delivered by recipients of medals and prizes noted in the following paragraphs. Three meetings, held at the Royal Institution, were the occasions of special lectures, also noted below. The remaining two meetings, in the rooms of the Royal Society, were devoted to lectures by distinguished visitors, Dr. A. J. Guinier of the Laboratoire . d'Essais, Paris, on "X-ray investigations of crystal lattice imperfections" (*Proceedings*, 57 (1945), p. 310), and Prof. B. Edlén of the University of Lund, on "The spectra of the inert gases ".

At the Royal Institution on 23 May, Prof. E. N. da C. Andrade delivered his Presidential Address on " The history and future of the Physical Society " (*Proceedings*, 57 (November 1945), p. xxi), and an Annual General Meeting followed for the presentation of the Reports of the Council and the Honorary Treasurer for 1944 and for the election of Officers and Council for 1945–46.

The Council records its thanks to the President and Council of The Royal Society, the Managers of the Royal Institution, the Director of the Science Museum and the Governing Body of Imperial College for their hospitality.

THOMAS YOUNG ORATION

At the Royal Institution on 29 June Professor Ragnar Granit (Stockholm) delivered the fourteenth Oration, the subject of which was " The electrophysiological analysis of the fundamental problem of colour reception " (*Proceedings*, 57 (1945), p. 447).

GUTHRIE LECTURE

The twenty-ninth Guthrie Lecture was delivered at the Royal Institution on 6 July by Professor Arturo Duperier (University of Madrid and Imperial College, London), who took as his subject " The Geophysical aspect of cosmic rays " (*Proceedings*, 57 (1945), p. 464).

CHARLES CHREE MEDAL AND PRIZE

At the Science Museum on 6 December the third (1945) Medal and Prize were presented to Dr. John A. Fleming, Director of the Department of Terrestrial Magnetism of the Carnegie Institution, Washington, the award being in recognition of Dr. Fleming's long and valuable service to Geophysics. Dr. Fleming afterwards delivered the third Charles Chree Address, the subject of which was " Geomagnetic secular variations and surveys " (*Proceedings*, 58 (1946), p. 213).

DUDDELL MEDAL

The twenty-first (1944) Medal was presented to the late Dr. F. W. Aston at a Science Meeting at the Royal Institution on 23 May, the presentation having been postponed on account of the medallist's ill-health. The Council records with great regret the death of Dr. Aston on 20 November 1945.

The Council awarded the twenty-second (1945) Medal to Professor J. T. Randall (University of St. Andrews) for his work on phosphors and the cavity magnetron. The presentation took place at the Science Museum on 12 December and was followed by a lecture and demonstrations by the medallist (*Proceedings*, 58 (1946), p. 247).

CHARLES VERNON BOYS PRIZE

The Council made the first (1945) award of the Prize to Dr. A. H. S. Holbourn (Clarendon Laboratory, Oxford) for his measurement of the angular momentum of radiation. A discourse on this work was given by the prizeman on the occasion of the presentation at the Science Museum on 12 December.

ADDENBROOKE BEQUEST

The volumes selected for the Society's Library in 1945 have not yet been received. The titles will be announced in due course in the Notices to Members and communicated to appropriate bodies.

PROCEEDINGS

A slightly bigger official allowance of paper made possible a small increase both in the size, and also in the number of copies printed, of Volume 57 (1945), as compared with the previous volume. The space devoted to advertisements in each of the six Parts remained unchanged. The Council again records its thanks to Dr. A. F. C. Pollard for his valuable services in the preparation of the U.D.C. Index Slips, which are supplied to

those members and subscribers who apply for them. The reprinting of those war-time issues the stocks of which have been exhausted is being continued as the supply of paper permits, but progress is still disappointingly slow.

REPORTS ON PROGRESS IN PHYSICS

It is regretted that very slow progress has been made in the printing of Volume 10 (1944–45), which it was hoped to publish early in 1946, and in the reprinting of Volume 4 (1937), which should have been ready during 1945. Steps are being taken to expedite the appearance of both these volumes. Copies of all the available volumes (only Vols. 7, 8 and 9 remain at the time of writing this Report) have been in great demand, especially since the liberation of European countries from former enemy occupation. The continued co-operation of the American Institute of Physics in supplying copies of the *Reports* to American physicists is again most gratefully acknowledged.

REPRESENTATION ON OTHER BODIES

The following were the Society's representatives on other Bodies in 1945 :—

National Committee for Physics : Dr. Ezer Griffiths, Prof. N. F. Mott, Mr. R. S. Whipple.
National Committee for Scientific Radio : Mr. J. A. Ratcliffe, Sir Robert Watson-Watt.
Committee of Management of Science Abstracts : Mr. J. H. Awbery, Prof. A. Ferguson, Dr. W. Jevons, Dr. D. Owen.
Board of the Institute of Physics : Prof. A. O. Rankine, Dr. W. Jevons.
Council of the British Society for International Bibliography : Prof. L. C. Martin.
Parliamentary and Scientific Committee : Prof. E. N. da C. Andrade, Prof. A. O. Rankine.
Joint Committee to facilitate Co-operation between the Physical Society and the Institute of Physics : Prof. E. N. da C. Andrade, Prof. D. Brunt, Prof. A. O. Rankine.
Joint Publications Board of the Physical Society and the Institute of Physics : Mr. J. H. Awbery, Prof. G. I. Finch, Dr. W. S. Stiles, Dr. W. D. Wright.
R. W. Paul Instrument Fund Committee : Prof. A. O. Rankine.
Röntgen Jubilee Celebrations Committee : Dr. W. Jevons.
British Standards Institution Committees : Prof. H. T. Flint.
International Psychrometry Committee : Prof. D. Brunt.
Association of Scientific Workers Conference Committee : Dr. W. B. Mann.
Royal Institute of Chemistry Conference on Definitions of Scientific Terms : Dr. D. Owen, Prof. A. O. Rankine, Dr. H. Shaw, Dr. W. D. Wright.

OBITUARY

The Council records with regret the deaths of the following Fellows :—Professor Eugène Bloch, Mr. W. M. Brett, Mr. J. E. Calthrop, Professor J. K. Catterson-Smith, Mr. W. D. Eggar, Professor Charles Fabry (Honorary Fellow), Mr. G. S. Fawcett, Sir Ambrose Fleming (Original Fellow), Mr. A. T. Hare, Mr. G. A. Heath, Major E. O. Henrici, Mr. D. E. Jones, Mr. W. Lucas, Major C. E. S. Phillips, Sir Napier Shaw, Dr. R. F. Schmid, Dr. J. F. Sirks, Professor E. J. Williams, Mr. J. Young.

MEMBERSHIP

Despite heavy losses by death, there was a notable expansion of total membership for the fifth year in succession ; it has been exceeded on only two previous occasions : in 1932, when members of the Optical Society came in, and in 1942, when there was a very large influx of Student Members. The number of transfers from Student Membership to Fellowship in 1945 was more than double that in any previous year ; the number of newly elected Fellows was exceeded only in 1942 ; and the net increase of Fellowship was far greater than in any year excepting 1932. There is a slight net decrease in the Student Membership, in spite of the fact that the number elected during the year was exceeded only in 1942 and in 1943.

Roll of Membership	Hon. Fellows	Hon. Fellow, Optical Society*	Ex-officio Fellows	Fellows	Student Members	Total
Totals, 31 Dec. 1944	11	1	4	1037	279	1382
Changes during 1945 Newly elected				56 ⎱108	68	
Transferred				52 ⎰	52 ⎱	
Deceased	1			19 ⎱ 28	5 ⎰72	
Resigned				9 ⎰	15 ⎱	
Lapsed or suspended						
Net increase	−1			80	−4	75
Totals, 31 Dec. 1945	10	1	4	1167	275	1457

* Excluding one who is also an Honorary Fellow of the Physical Society (in previous column).

GROUPS

Colour Group

The fifth Annual General Meeting of the Group was held at Imperial College on 21 March 1945, when Mr. J. Guild and Dr. W. D. Wright were re-elected as Chairman and as Honorary Secretary, respectively, and the Committee for 1945–46 was elected.

During the year six Science Meetings were held, particulars of which will be found elsewhere in the *Proceedings* (57, (1945) pp. ix, x and xi; and 58 (1946), pp. vi, vii and viii).

Considerable progress was made with the reports of the Sub-Committees on Defective Colour Vision in Industry and Colour Terminology, and the former should be ready for publication in 1946.

Optical Group

At the fourth Annual General Meeting of the Group, which was held at Imperial College on 20 April 1945, Instr.-Capt. T. Y. Baker and Mr. E. W. H. Selwyn were elected as Chairman and as Honorary Secretary, respectively, and the Committee for 1945–46 was elected.

Five Science Meetings were held in 1945, and are briefly recorded elsewhere in the *Proceedings* (57 (1945), pp. ix, x and xii; and 58 (1946), pp. vi and vii). Three of the papers read at these meetings have been published in the *Proceedings* (57 (1945), pp. 286, 403 and 543).

Low-Temperature Group

The Low-Temperature Group was inaugurated at a General Meeting held in the rooms of the Royal Society on 17 October 1945, when the Officers and Committee for 1945–46 were elected, Sir Alfred Egerton being elected as Chairman, Sir Charles Darwin as Vice-Chairman, and Mr. J. H. Awbery as Honorary Secretary.

At the first Science Meeting, also held on 17 October, Mr. A. M. Clark delivered a lecture on " Some technical problems of gas separation ", which is to be printed in the *Chemical Trade Journal.*

Membership

The membership of the three Groups on 31 December 1945 was as follows :—

	Colour Group	Optical Group	Low-Temperature Group
Members of the Physical Society	87	156	30
Members of participating bodies	63	68	22
Members of firms subscribing for sustaining membership	18	97	..
Other Members	10	5	2
Totals	178	326	54

REPORT OF THE HONORARY TREASURER
FOR THE YEAR ENDED 31 DECEMBER 1945

Three recommendations by the auditors have been embodied in the Accounts for 1945 : (i) the balance of £83 1s., which has remained in the Progress Reports Reserve Account for the last ten years and is no longer likely to be called upon, is now transferred to the general fund ; (ii) a grant of £75 is made from the general fund to the Duddell Memorial Trust Fund in order to meet the present deficit ; and (iii) the expenditure of £154 for office furniture and equipment in 1945 is charged, as was that in 1938, against the Herbert Spencer Fund, the balance of which would normally have been exhausted several years ago in accordance with the terms of the bequest.

With these adjustments the accounts show an excess of income over expenditure of £869, which is £275 less than in 1944 ; the total income (£5319) and the total expenditure (£4950) were higher by £469 and £744, respectively.

Three of the main sources of income (subscriptions, advertisements and dividends) showed welcome increments, while the fourth (sales of publications) showed little net change, the drop on sales of the *Reports* being nearly made up by the large rise on sales of the *Proceedings* and special publications.

A further grant (£450 as against £300) by the Royal Society from the Rockefeller Gift is gratefully acknowledged, and will be expended on reprinting some of the war-time Parts of the *Proceedings* as soon as conditions allow.

Administration expenses were about £238 higher than in 1944, the fall of expenditure due to the cessation of fire-watching at the office premises being outweighed by the not unexpected rise in secretarial and clerical expenses.

Expenditure on the Society's publications fell by over £300, in spite of the rise in cost of producing the *Proceedings*, because no volume of the *Reports on Progress in Physics* appeared during the year. An important increase in expenditure is that in respect of *Science Abstracts*.

The total value of the investments increased during the year by £4498 to £15,764. The Addenbrooke Bequest and the donations to the Holweck Prize Fund account for £1212 of this increase, and there was a new investment of £1000 in 3 % Savings Bonds ; the remaining £2286 represents the appreciation of the older investments. (The values of these shown in the Balance Sheet are those obtaining on 31 December 1939.) The investment of the capital of the Holweck Prize in 3 % Defence Bonds will enable an annual award of £100 to be maintained for ten years.

<div align="right">

(Signed) CLIFFORD C. PATERSON,
Honorary Treasurer.

</div>

12 *April* 1946.

SPECIAL FUNDS
W. F. STANLEY TRUST FUND

	£ s. d.		£ s.
d to Balance Sheet . . .	259 0 0	£300 Southern Railway Preferred Ordinary Stock	199
		£442 Southern Railway Deferred Ordinary Stock	60
	£259 0 0		£259

DUDDELL MEMORIAL TRUST FUND
CAPITAL

	£ s. d.		£
d to Balance Sheet . . .	374 0 0	£400 3¼% War Loan Inscribed "B" Account	374

REVENUE

	£ s. d.		£
:e on 31 December 1944 . .	27 17 3	Interest on War Loan	14
·aria	46 5 0	Grant from general fund . .	75
's and Certificates . . .	14 2 0		
:e carried to Balance Sheet . .	15 9		
	£89 0 0		£89

SPECIAL FUNDS (*contd.*)

HERBERT SPENCER LEGACY

	£	s.	d.		£
ıditure on Furniture and Equipment				Balance on 31st December 1944 . .	239
ing the year . . .	154	3	3		
ce carried to Balance Sheet . .	85	4	3		
	£239	7	6		£239

CHARLES CHREE MEDAL AND PRIZE FUND
CAPITAL

	£	s.	d.		£
ce carried to Balance Sheet . .	1865	16	4	Balance on 31 December 1944 . .	1865

REVENUE

	£	s.	d.		£
. . . . . .	131	5	0	Balance on 31 December 1944 . .	83
l and Certificate . . .	7	19	0	Interest on Investments . . .	68
ce carried to Balance Sheet . .	12	19	10		
	£152	3	10		£152

CHARLES VERNON BOYS PRIZE FUND
CAPITAL

	£	s.	d.		£
ce carried to Balance Sheet . .	900	0	0	£1132 16s. 10d. 2½% Consols "B" account	900

REVENUE

	£	s.	d.		£
. . . .	26	5	0	Balance on 31 December 1944 . .	?
ficate	26	10	0	Interest on Investment . . .	28
				Balance carried to Balance Sheet . .	17
	£52	15	0		£52

HOLWECK PRIZE FUND

	£	s.	d.		£
ıce carried to Balance Sheet . . .	887	4	11	Donations received	88?

ADDENBROOKE BEQUEST
CAPITAL

	£	s.	d.		£
ıce carried to Balance Sheet . .	337	0	0	£384 6s. 7d. 2½% Consols "D" account	33?

REVENUE

	£	s.	d.		
ıce carried to Balance Sheet . .	9	12	0	Interest on Investment	

LIFE COMPOSITION FUND ON 31 DECEMBER 1945

	£	s.	d.
19 Fellows paid £10	190	0	0
1 Fellow paid £15	15	0	0
15 Fellows paid £21	315 .	0	0
44 Fellows paid £31 10s.	1386	0	0
	£1906	0	0

BALANCE SHEET AS ON 31 DECEMBER 1945

LIABILITIES

	£ s. d.	£ s. d.
Sundry Creditors		567 3 10
Life Compositions:		
As on 31 December 1944	1674 10 0	
Add Payments during year	252 0 0	
	1926 10 0	
Less Transfer to Income and Expenditure Account	20 10 0	1906 0 0
Subscriptions received in advance:		
Members	35 13 6	
Publications	404 1 11	439 15 5
Special Funds:		
W. F. Stanley Trust Fund	259 0 0	
Duddell Memorial Trust Fund	374 15 9	
Herbert Spencer Legacy	85 4 3	
Charles Chree Medal and Prize Fund	1878 16 2	
Charles Vernon Boys Prize Fund	882 12 11	
Holweck Prize Fund	887 4 11	
Addenbrooke Bequest	346 12 0	4714 6 0
General Fund:		
As on 31 December 1944	6805 7 8	
Add "Progress Reports" Reserve Account	83 1 0	
,, Balance of Income and Expenditure Account	869 1 10	7757 10 6
		£15384 15 9

ASSETS

	£ s. d.	£ s. d.	£ s. d.
Investments:			
W. F. Stanley Trust Fund	259 0 0		
Duddell Memorial Fund	374 0 0		
Charles Chree Medal and Prize Fund	1866 10 0		
Charles Vernon Boys Prize Fund	900 0 0		
Holweck Prize Fund	875 0 0		
Addenbrooke Bequest	337 0 0		
General Fund	7505 10 0	12117 0 0	
(Market Value on 31 December 1945: £15764)			
Dividends due on Investments		74 5 7	
Inland Revenue—Income Tax recoverable for 1944 and 1945		141 15 10	
Subscriptions due		25 7 0	
Sundry Debtors		293 1 5	
Stock of Paper and Binding Material		173 0 0	
30th Exhibition (1946):—Expenditure to date		640 9 3	
Cash in Post Office Savings Bank Account	1940 11 7		
,, in hand	24 15 4	1965 6 11	
Less overdraft on current account		45 10 3	1919 16 8
			£15384 15 9

We have audited the above Balance Sheet and have obtained all the information and explanations we have required. We have verified the bank balances and the Investments. In our opinion such Balance Sheet is properly drawn up so as to exhibit a true and correct view of the state of the Society's affairs according to the best of our information and the explanations given to us and as shown by the books of the Society.

SPENCER HOUSE, SOUTH PLACE, E.C. 2.
5th April 1946

KNOX, CROPPER & Co.,
Chartered Accountants.

INCOME AND EXPENDITURE ACCOUNT FOR THE YEAR ENDED 31 DECEMBER 1945

1944 £	EXPENDITURE	£ s. d.	£ s. d.
	To		
225	"Science Abstracts"		801 0 0
1686	"Proceedings"	2104 12 3	
59	Agenda Papers and Notices	93 1 0	
779	"Reports on Progress in Physics"	21 0 0	
			2218 13 3
234	Postage on Publications and General Correspondence		335 8 8
62	Expenses at Meetings		94 18 9
26	Honoraria to Special Lecturers		52 10 0
	Administration Expenses:		
910	Secretarial and Clerical Assistance and Office Expenses	1287 18 8	
295	1 Lowther Gardens	84 18 9	
			1372 17 5
	Grant to Duddell Memorial Fund		75 0 0
1144	Balance, being excess of Income over Expenditure, carried forward to Accumulated Fund		869 1 10
£5350			£5819 9 11

1944 £	INCOME	£ s. d.	£ s. d.	£ s. d.
	By			
	Subscriptions:			
54	Entrance Fees	77 14 0		
2005	Fellows	2175 1 6		
147	Student Members	143 16 0		
23	Colour Group	28 10 0		
66	Optical Group	50 7 0		
148	For "Science Abstracts" and Advance Proofs	142 7 0		2617 15 6
	Sales:			
1019	"Proceedings"	1328 1 9		
852	"Reports on Progress in Physics"	495 4 0		
48	Special Publications	88 9 4		1911 15 1
334	Advertisements in "Proceedings"			448 4 8
304	Dividends from Investments			371 4 8
50	Transfer from Life Composition Fund of amounts paid by Fellows now deceased			20 10 0
300	Publications Grant by the Royal Society from the Rockefeller Foundation Gift			450 0 0
£5350				£5819 9 11

COLOURS OF TOTAL RADIATORS EXPRESSED ON THE C.I.E. TRICHROMATIC SYSTEM FOR THE TEMPERATURE RANGE 0-1-660 MIREDS ($C_2 = 14384·8$)

By H. G. W. HARDING,

National Physical Laboratory, Teddington

ABSTRACT. Six-figure tables of the trichromatic coefficients of the colours of Planckian (total or black-body) radiators expressed on the C.I.E. trichromatic system (1931) are given for the temperature range 0-1-660 mireds. The colours refer to energy distributions calculated from Planck's formula with a value 14384·8 micron degrees for the constant C_2. Additional tables give the differences in the coefficients of the colours from the tabulated values for a change in the value of C_2 from 14384·8 to 14320 and from 14384·8 to 14350. These additional tables can be used to calculate the colours if other values of C_2 are used in Planck's formula.

TABLES which give the trichromatic coefficients of the colours of total radiators expressed on the C.I.E. system (1931) for the temperature range 1500–10–9000° K. have been published (Harding, 1944). These six-figure tables enable the coefficients of the colours of total radiators at any temperature within the given range to be calculated to five-figure accuracy by linear interpolation. To extend the range of these tables to cover the temperature range between 9000° K. and infinite temperature, the colours have been recalculated on a mired temperature scale and the range extended to infinite temperature. A mired, or micro-reciprocal-degree, is a million divided by the temperature of the radiator expressed in °K. (Priest, 1933).

The colours of the total radiators for the range 0-1-660 mireds were obtained by calculating the coefficients of the colours first for 0–10–660 mireds, then interpolating these coefficients to tenths.

The colours of the total radiators for 0–10–110 mireds had not been published and so they were calculated in the following way. The spectral distributions of the energies of the radiators were calculated from Planck's formula,

$$E_{\lambda, M}\, d\lambda = C_1 \lambda^{-5} \left[\exp \frac{C_2 M}{\lambda\, 10^6} - 1 \right]^{-1} d\lambda,$$

where $E_{\lambda, M}\, d\lambda$ represents the amount of energy radiated at M mireds between the wave-lengths $\lambda \pm \dfrac{d\lambda}{2}$ (microns) and C_1 and C_2 are constants. Since the energy values were made equal to 100·0000 at 0·56 μ, the value of C_1 was not required. The value 14384·8 micron degrees was given to C_2 (Birge, 1941). The distributions of energy were checked by fifth differences and were estimated to be correct to about one part in a million. The trichromatic coefficients of the colours of the radiators were then calculated from the spectral distributions of energy, using the

Condensed Tables for Colour Computation for the equal-energy stimulus (Smith, ·1934).

The colours of the radiators at 120–10–660 mireds were calculated from the previously published tables and the trichromatic coefficients were altered in some instances by a unit in the sixth decimal place to make the fifth differences smooth.

The trichromatic coefficients obtained by the methods just explained for the range 0–10–660 mireds were then interpolated to tenths, giving values at 0–1–660 mireds. Alterations of a unit in the sixth decimal place were made to the Z coefficients where necessary to make the sum of all three coefficients exactly unity. These coefficients are given in table 1.

The alterations to be made to the trichromatic coefficients if C_2 is changed from 14384·8 to 14320 were calculated for the mired range 0–10–660 mireds in the following way. Two illuminants have the same spectral distribution of energy, consequently the same colour, if the value of C_2M in Planck's formula remains unaltered. It follows that the trichromatic coefficients of the colour of a total radiator at M_{14320} mireds, calculated with $C_2 = 14320$, will be the same as those of a radiator at 14320 $(14384·8)^{-1} M_{14384·8} = 0·9954952 M_{14384·8}$ mireds if 14384·8 has been used for the colour calculation. The differences between the trichromatic coefficients for $M_{14384·8}$ mireds and $0·9954952 M_{14384·8}$ mireds were therefore calculated from the trichromatic coefficients in table 1 and are tabulated in table 2. If the value of C_2 is 14350, the alterations to the coefficients in table 1 are given by the differences between the coefficients for $M_{14384·8}$ mireds and $14350 (14384·8)^{-1} M_{14384·8} = 0·9975808 M_{14384·8}$ mireds. These differences are tabulated in table 3. All the differences in tables 2 and 3 were differenced to the third order and an alteration of a unit in the sixth decimal place was made to a few values to make the third differences smooth.

For other likely values of C_2 different from 14320 or 14350, linear interpolation of the values given either in table 2 or table 3 gives the differences with errors less than one or two units in the sixth decimal place; for instance, if $C_2 = 14360$ the X coefficient corresponding to 150 mireds is equal to

$$0·310998 - 0·000237 \frac{14384·8 - 14360}{14384·8 - 14350} = 0·310998 - 0·000169 = 0·31083.$$

The six figures for the trichromatic coefficients are only quoted here in order that the smoothed tables might serve as a basis for the preparation of other tables that might be required with a value of C_2 other than 14384·8.

The values quoted in tables 1, 2 and 3 might in a few instances be one or two units in the sixth decimal place different from those in the previously published tables. These differences were considered to be necessary to smooth the values in the present tables.

REFERENCES

BIRGE, R. T., 1941. *Rev. Mod. Phys.* **13**, 237.
HARDING, H. G. W., 1944. *Proc. Phys. Soc.* **56**, 305.
PRIEST, I. G., 1933. *J. Opt. Soc. Amer.* **23**, 41.
SMITH, T., 1934. *Proc. Phys. Soc.* **46**, 372.

Table 1

Mireds $(T \, °\text{K.})^{-1} \times 10^6$ $(C_2 = 14384 \cdot 8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
0	239906	234131	525963
1	240163.	234513	525324
2	240423	234898	524679
3	240687	235287	524026
4	240953	235680	523367
5	241221	236077	522702
6	241493	236477	522030
7	241768	236882	521350
8	242046	237290	520664
9	242327	237702	519971
10	242610	238118	519272
11	242897	238538	518565
12	243187	238961	517852
13	243479	239388	517133
14	243775	239819	516406
15	244074	240254	515672
16	244376	240692	514932
17	244680	241134	514186
18	244988	241580	513432
19	245299	242030	512671
20	245613	242483	511904
21	245930	242940	511130
22	246250	243401	510349
23	246574	243865	509561
24	246900	244333	508767
25	247229	244805	507966
26	247562	245280	507158
27	247897	245759	506344
28	248236	246241	505523
29	248578	246727	504695
30	248923	247216	503861
31	249271	247709	503020
32	249622	248206	502172
33	249977	248705	501318
34	250334	249209	500457
35	250695	249715	499590
36	251058	250225	498717
37	251425	250739	497836
38	251795	251255	496950
39	252168	251775	496057
40	252545	252299	495156
41	252924	252825	494251
42	253307	253355	493338
43	253692	253888	492420
44	254081	254424	491495

H. G. W. Harding

Table 1 (continued)

Mireds $(T\,^{\circ}\text{K.})^{-1} \times 10^6$ $(C_2 = 14384 \cdot 8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
45	254473	254963	490564
46	254868	255505	489627
47	255266	256050	488684
48	255668	256598	487734
49	256072	257150	486778
50	256480	257704	485816
51	256890	258261	484849
52	257304	258821	483875
53	257721	259384	482895
54	258141	259950	481909
55	258564	260518	480918
56	258990	261090	479920
57	259419	261664	478917
58	259851	262240	477909
59	260287	262820	476893
60	260725	263402	475873
61	261166	263986	474848
62	261611	264573	473816
63	262058	265162	472780
64	262509	265754	471737
65	262962	266349	470689
66	263419	266945	469636
67	263879	267544	468577
68	264341	268145	467514
69	264807	268749	466444
70	265275	269355	465370
71	265747	269962	464291
72	266221	270572	463207
73	266698	271184	462118
74	267178	271798	461024
75	267662	272414	459924
76	268148	273032	458820
77	268636	273652	457712
78	269128	274274	456598
79	269623	274898	455479
80	270120	275523	454357
81	270621	276150	453229
82	271124	276779	452097
83	271630	277409	450961
84	272138	278041	449821
85	272650	278675	448675
86	273164	279310	447526
87	273681	279946	446373
88	274201	280584	445215
89	274723	281224	444053

Table 1 (*continued*)

Mireds $(T\,°\text{K.})^{-1}\times 10^6$ $(C_2=14384\cdot 8)$	Trichromatic coefficients		
	$x\times 10^6$	$y\times 10^6$	$z\times 10^6$
90	275249	281864	442887
91	275776	282506	441718
92	276307	283150	440543
93	276840	283794	439366
94	277376	284440	438184
95	277914	285086	437000
96	278455	285734	435811
97	278999	286383	434618
98	279545	287033	433422
99	280093	287684	432223
100	280644	288336	431020
101	281197	288989	429814
102	281754	289642	428604
103	282313	290296	427391
104	282874	290951	426175
105	283437	291607	424956
106	284003	292263	423734
107	284572	292920	422508
108	285143	293578	421279
109	285716	294236	420048
110	286291	294894	418815
111	286869	295553	417578
112	287449	296212	416339
113	288031	296872	415097
114	288615	297533	413852
115	289202	298193	412605
116	289791	298853	411356
117	290382	299514	410104
118	290976	300175	408849
119	291571	300836	407593
120	292169	301497	406334
121	292769	302158	405073
122	293371	302819	403810
123	293975	303480	402545
124	294581	304141	401278
125	295190	304802	400008
126	295800	305463	398737
127	296412	306123	397465
128	297027	306783	396190
129	297643	307443	394914
130	298261	308103	393636
131	298881	308762	392357
132	299503	309421	391076
133	300127	310080	389793

Table 1 (*continued*)

Mireds $(T^\circ\text{K.})^{-1} \times 10^6$ $(C_2 = 14384\cdot8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
135	301379	311396	387225
136	302009	312053	385938
137	302640	312709	384651
138	303273	313365	383362
139	303908	314020	382072
140	304544	314675	380781
141	305182	315329	379489
142	305822	315982	378196
143	306463	316634	376903
144	307106	317286	375608
145	307751	317937	374312
146	308397	318587	373016
147	309045	319236	371719
148	309695	319884	370421
149	310346	320531	369123
150	310998	321177	367825
151	311652	321822	366526
152	312307	322466	365227
153	312964	323109	363927
154	313623	323751	362626
155	314282	324392	361326
156	314943	325031	360026
157	315606	325669	358725
158	316270	326307	357423
159	316935	326942	356123
160	317601	327577	354822
161	318269	328210	353521
162	318937	328842	352221
163	319608	329473	350919
164	320279	330102	349619
165	320951	330730	348319
166	321625	331356	347019
167	322300	331980	345720
168	322976	332604	344420
169	323653	333226	343121
170	324331	333846	341823
171	325010	334465	340525
172	325691	335082	339227
173	326372	335697	337931
174	327054	336311	336635
175	327738	336923	335339
176	328422	337534	334044
177	329108	338142	332750
178	329794	338749	331457
179	330491	339355	330164

Table 1 (*continued*)

Mireds $(T^\circ\text{к.})^{-1}\times 10^6$ $(C_2=14384\cdot 8)$	Trichromatic coefficients		
	$x\times 10^6$	$y\times 10^6$	$z\times 10^6$
180	331169	339958	328873
181	331858	340560	327582
182	332548	341159	326293
183	333238	341757	325005
184	333929	342353	323718
185	334621	342947	322432
186	335314	343539	321147
187	336008	344129	319863
188	336702	344717	318581
189	337397	345304	317299
190	338093	345888	316019
191	338789	346470	314741
192	339486	347050	313464
193	340184	347629	312187
194	340882	348205	310913
195	341581	348779	309640
196	342281	349351	308368
197	342981	349920	307099
198	343681	350488	305831
199	344382	351054	304564
200	345084	351617	303299
201	345786	352178	302036
202	346489	352737	300774
203	347192	353294	299514
204	347895	353849	298256
205	348599	354401	297000
206	349303	354951	295746
207	350007	355499	294494
208	350712	356045	293243
209	351418	356588	291994
210	352123	357129	290748
211	352829	357668	289503
212	353535	358204	288261
213	354241	358738	287021
214	354947	359270	285783
215	355654	359799	284547
216	356361	360326	283313
217	357068	360851	282081
218	357775	361373	280852
219	358482	361893	279625
220	359190	362410	278400
221	359898	362925	277177
222	360606	363437	275957
223	361314	363947	274739
224	362022	364454	273524

H. G. W. Harding

Table 1 (*continued*)

Mireds $(T°\text{к.})^{-1} \times 10^6$ $(C_2 = 14384 \cdot 8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
225	362730	364959	272311
226	363438	365461	271101
227	364146	365961	269893
228	364854	366458	268688
229	365563	366953	267484
230	366271	367445	266284
231	366979	367935	265086
232	367688	368422	263890
233	368396	368906	262698
234	369104	369389	261507
235	369812	369868	260320
236	370520	370345	259135
237	371228	370819	257953
238	371935	371291	256774
239	372643	371760	255597
240	373350	372227	254423
241	374057	372691	253252
242	374764	373152	252084
243	375470	373611	250919
244	376177	374067	249756
245	376883	374521	248596
246	377589	374972	247439
247	378294	375420	246286
248	378999	375866	245135
249	379704	376309	243987
250	380409	376749	242842
251	381113	377187	241700
252	381817	377622	240561
253	382521	378054	239425
254	383224	378483	238293
255	383927	378910	237163
256	384630	379335	236035
257	385332	379756	234912
258	386034	380175	233791
259	386735	380591	232674
260	387436	381005	231559
261	388136	381416	230448
262	388836	381824	229340
263	389535	382230	228235
264	390234	382632	227134
265	390933	383033	226034
266	391630	383430	224930
267	392328	383825	223847
268	393024	384217	222759
269	393720	384606	221674

Table 1 (*continued*)

Mireds $(T°\text{K.})^{-1} \times 10^6$ $(C_2 = 14384 \cdot 8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
270	394416	384993	220591
271	395111	385377	219512
272	395805	385758	218437
273	396499	386137	217364
274	397193	386513	216294
275	397885	386886	215229
276	398577	387257	214166
277	399269	387625	213106
278	399959	387990	212051
279	400650	388352	210998
280	401339	388712	209949
281	402028	389069	208903
282	402716	389423	207861
283	403403	389775	206822
284	404090	390124	205786
285	404776	390470	204754
286	405461	390814	203725
287	406146	391154	202700
288	406830	391493	201677
289	407513	391828	200659
290	408195	392161	199644
291	408876	392491	198633
292	409557	392819	197624
293	410237	393144	196619
294	410915	393466	195619
295	411593	393786	194621
296	412271	394103	193626
297	412947	394417	192636
298	413623	394729	191648
299	414297	395038	190665
300	414971	395344	189685
301	415644	395648	188708
302	416316	395949	187735
303	416987	396248	186765
304	417657	396543	185800
305	418327	396837	184836
306	418995	397127	183878
307	419663	397415	182922
308	420330	397701	181969
309	420995	397984	181021
310	421660	398264	180076
311	422324	398542	179134
312	422987	398816	178197
313	423648	399089	177263
314	424309	399358	176333

Table 1 (*continued*)

Mireds $(T \, °\text{к.})^{-1} \times 10^6$ $(C_2 = 14384 \cdot 8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
315	424969	399626	175405
316	425628	399890	174482
317	426286	400152	173562
318	426943	400412	172645
319	427599	400669	171732
320	428254	400923	170823
321	428908	401175	169917
322	429561	401424	169015
323	430213	401671	168116
324	430863	401915	167222
325	431513	402157	166330
326	432162	402397	165441
327	432809	402633	164558
328	433456	402868	163676
329	434102	403100	162798
330	434746	403329	161925
331	435389	403556	161055
332	436032	403780	160188
333	436673	404002	159325
334	437313	404222	158465
335	437952	404439	157609
336	438590	404654	156756
337	439226	404866	155908
338	439862	405076	155062
339	440497	405283	154220
340	441130	405488	153382
341	441762	405691	152547
342	442394	405891	151715
343	443024	406089	150887
344	443653	406284	150063
345	444280	406477	149243
346	444907	406668	148425
347	445533	406856	147611
348	446157	407042	146801
349	446780	407226	145994
350	447402	407407	145191
351	448023	407586	144391
352	448642	407763	143595
353	449261	407937	142802
354	449878	408109	142013
355	450494	408279	141227
356	451109	408446	140445
357	451722	408611	139667
358	452335	408774	138891

Table 1 (*continued*)

Mireds $(T °K.)^{-1} \times 10^6$ $(C_2 = 14384 \cdot 8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
360	453556	409093	137351
361	454165	409249	136586
362	454772	409403	135825
363	455379	409555	135066
364	455984	409704	134312
365	456588	409851	133561
366	457191	409997	132812
367	457793	410139	132068
368	458393	410280	131327
369	458992	410419	130589
370	459590	410555	129855
371	460187	410689	129124
372	460782	410821	128397
373	461376	410951	127673
374	461969	411079	126952
375	462561	411205	126234
376	463152	411329	125519
377	463741	411450	124809
378	464329	411569	124102
379	464916	411687	123397
380	465501	411802	122697
381	466085	411915	122000
382	466668	412026	121306
383	467250	412135	120615
384	467830	412242	119928
385	468409	412347	119244
386	468987	412450	118563
387	469564	412550	117886
388	470139	412649	117212
389	470713	412746	116541
390	471286	412841	115873
391	471858	412934	115208
392	472428	413025	114547
393	472997	413114	113889
394	473565	413202	113233
395	474131	413287	112582
396	474697	413370	111933
397	475261	413452	111287
398	475823	413532	110645
399	476385	413609	110006
400	476945	413685	109370
401	477504	413759	108737
402	478062	413831	108107
403	478618	413901	107481
404	479173	413969	106858

Table 1 (*continued*)

Mireds $(T\,°\mathrm{K.})^{-1}\times10^6$ ($C_2=14384\cdot8$)	Trichromatic coefficients		
	$x\times10^6$	$y\times10^6$	$z\times10^6$
405	479727	414035	106238
406	480280	414100	105620
407	480831	414163	105006
408	481381	414224	104395
409	481930	414283	103787
410	482477	414340	103183
411	483023	414396	102581
412	483568	414449	101983
413	484112	414501	101387
414	484654	414552	100794
415	485195	414600	100205
416	485735	414647	99618
417	486273	414692	99035
418	486810	414735	98455
419	487346	414777	97877
420	487881	414817	97302
421	488414	414855	96731
422	488946	414892	96162
423	489477	414927	95596
424	490007	414960	95033
425	490535	414992	94473
426	491062	415022	93916
427	491588	415050	93362
428	492112	415077	92811
429	492635	415102	92263
430	493157	415126	91717
431	493678	415148	91174
432	494197	415168	90635
433	494715	415187	90098
434	495232	415204	89564
435	495748	415220	89032
436	496262	415234	88504
437	496775	415247	87978
438	497287	415258	87455
439	497798	415267	86935
440	498307	415275	86418
441	498815	415281	85904
442	499322	415286	85392
443	499827	415290	84883
444	500332	415292	84376
445	500835	415292	83873
446	501336	415291	83373
447	501837	415289	82874
448	502336	415285	82379

Table 1 (*continued*) .

Mireds $(T°\text{K.})^{-1} \times 10^6$ $(C_2 = 14384.8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
450	503331	415273	81396
451	503827	415265	80908
452	504321	415255	80424
453	504814	415244	79942
454.	505306	415232	79462
455	505796	415218	78986
456 .	506286	415203	78511
457	506774	415186	78040
458	507261	415169	77570
459	507747	415149	77104
460	508231	415129	76640
461	508714	415107	76179
462	509196	415084	75720
463	509677	415060	75263
464	510156	415034	74810
465	510635	415007	74358
466	511112	414979	73909
467	511588	414950	73462
468	512062	414919	73019
469	512536	414887	72577
470	513008	414854	72138
471	513479	414820	71701
472	513949	414784	71267
473	514418	414747	70835
474	514885	414709	70406
475	515351	414670	69979
476	515817	414630	69553
477	516280	414588	69132
478	516743	414545	68712
479	517205	414501	68294
480	517665	414456	67879
481	518124	414410	67466
482	518582	414362	67056
483	519039	414314	66647
484	519494	414264	66242
485	519949	414213	65838
486	520402	414161	65437
487	520854	414108	65038
488	521305	414054	64641
489	521754	413998	64248
490	522203	413942	63855
491	522650	413885	63465
492	523097	413826	63077
493.	523542	413767	62691
494	523986	413706	62308

H. G. W. Harding

Table 1 (*continued*)

Mireds $(T°\text{K.})^{-1} \times 10^6$ ($C_2 = 14384·8$)	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
495	524428	413644	61928
496	524870	413582	61548
497	525310	413518	61172
498	525750	413453	60797
499	526188	413388	60424
500	526625	413321	60054
501	527061	413253	59686
502	527496	413185	59319
503	527929	413115	58956
504	528362	413044	58594
505	528793	412973	58234
506	529223	412900	57877
507	529652	412826	57522
508	530080	412752	57168
509	530507	412676	56817
510	530933	412600	56467
511	531358	412523	56119
512	531781	412444	55775
513	532204	412365	55431
514	532625	412285	55090
515	533045	412204	54751
516	533464	412122	54414
517	533882	412039	54079
518	534299	411956	53745
519	534715	411871	53414
520	535130	411786	53084
521	535544	411700	52756
522	535956	411613	52431
523	536368	411525	52107
524	536778	411437	51785
525	537187	411347	51466
526	537596	411257	51147
527	538003	411166	50831
528	538409	411074	50517
529	538814	410981	50205
530	539218	410888	49894
531	539621	410794	49585
532	540023	410699	49278
533	540424	410603	48973
534	540824	410507	48669
535	541223	410409	48368
536	541620	410311	48069
537	542017	410213	47770
538	542413	410113	47474

Table 1 (*continued*)

Mireds $(T\ °\text{K.})^{-1}\times10^6$ $(C_2=14384\cdot8)$	Trichromatic coefficients		
	$x\times10^6$	$y\times10^6$	$z\times10^6$
540	543201	409912	46887
541	543594	409810	46596
542	543985	409708	46307
543	544376	409605	46019
544	544765	409501	45734
545	545154	409397	45449
546	545541	409292	45167
547	545928	409186	44886
548	546313	409080	44607
549	546698	408973	44329
550	547081	408865	44054
551	547463	408757	43780
552	547845	408647	43508
553	548225	408538	43237
554	548604	408427	42969
555	548983	408316	42701
556	549360	408205	42435
557	549737	408092	42171
558	550112	407980	41908
559	550487	407866	41647
560	550860	407752	41388
561	551233	407637	41130
562	551604	407522	40874
563	551975	407406	40619
564	552344	407290	40366
565	552713	407173	40114
566	553081	407056	39863
567	553447	406937	39616
568	553813	406819	39368
569	554178	406700	39122
570	554542	406580	38878
571	554905	406460	38635
572	555267	406339	38394
573	555628	406218	38154
574	555988	406096	37916
575	556348	405973	37679
576	556706	405850	37444
577	557063	405727	37210
578	557420	405603	36977
579	557775	405479	36746
580	558130	405354	36516
581	558484	405229	36287
582	558836	405103	36061
583	559188	404977	35835
584	559539	404850	35611

Table 1 (*continued*)

Mireds $(T°\text{K.})^{-1} \times 10^6$ $(C_2 = 14384 \cdot 8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
585	559889	404723	35388
586	560239	404595	35166
587	560587	404467	34946
588	560934	404338	34728
589	561280	404209	34511
590	561626	404080	34294
591	561971	403950	34079
592	562314	403820	33866
593	562657	403689	33654
594	562999	403558	33443
595	563340	403426	33234
596	563680	403294	33026
597	564019	403162	32819
598	564358	403029	32613
599	564695	402896	32409
600	565032	402762	32206
601	565368	402628	32004
602	565703	402494	31803
603	566037	402359	31604
604	566370	402224	31406
605	566703	402088	31209
606	567035	401952	31013
607	567366	401816	30818
608	567696	401679	30625
609	568025	401542	30433
610	568353	401405	30242
611	568681	401267	30052
612	569007	401129	29864
613	569333	400991	29676
614	569658	400852	29490
615	569982	400713	29305
616	570306	400573	29121
617	570628	400433	28939
618	570950	400293	28757
619	571271	400153	28576
620	571591	400012	28397
621	571910	399871	28219
622	572229	399730	28041
623	572547	399588	27865
624	572863	399446	27691
625	573180	399304	27516
626	573495	399161	27344
627	573809	399018	27173
628	574123	398875	27002
629	574436	398732	26832

Table 1 (*contined*)

Mireds $(T°\text{K.})^{-1} \times 10^6$ $(C_2 = 14384·8)$	Trichromatic coefficients		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
630	574748	398588	26664
631	575059	398444	26497
632	575370	398300	26330
633	575680	398156	26164
634	575989	398011	26000
635	576297	397866	25837
636	576604	397721	25675
637	576911	397576	25513
638	577217	397430	25353
639	577522	397284	25194
640	577826	397138	25036
641	578130	396992	24878
642	578432	396845	24723
643	578734	396698	24568
644	579036	396551	24413
645	579336	396404	24260
646	579636	396256	24108
647	579935	396109	23956
648	580234	395961	23805
649	580531	395812	23657
650	580828	395664	23508
651	581124	395515	23361
652	581420	395367	23213
653	581714	395218	23068
654	582008	395068	22924
655	582301	394919	22780
656	582594	394769	22637
657	582886	394620	22494
658	583177	394470	22353
659	583467	394319	22214
660	583757	394169	22074

Table 2

Mireds $(T°\text{K.})^{-1} \times 10^6$ $(C_2 = 14384·8)$	Change in trichromatic coefficients if C_2 is altered from 14384·8 to 14320		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
0	− 0	− 0	+ 0
10	13	19	32
20	28	41	69
30	47	66	113
40	68	94	162

Table 2 (*continued*)

Mireds $(T\ °\text{к.})^{-1}\times10^6$ ($C_2=14384\cdot8$)	Change in trichromatic coefficients if C_2 is altered from 14384·8 to 14320		
	$x\times10^6$	$y\times10^6$	$z\times10^6$
50	− 92	−125	+217
60	119	157	276
70	148	191	339
80	179	225	404
90	213	260	473
100	248	294	542
110	285	326	611
120	323	357	680
130	362	386	748
140	401	413	814
150	441	436	877
160	480	458	938
170	519	475	994
180	558	489	1047
190	595	500	1095
200	632	508	1140
210	667	512	1179
220	701	513	1214
230	734	510	1244
240	765	505	1270
250	794	496	1290
260	821	485	1306
270	846	471	1317
280	870	454	1324
290	891	436	1327
300	911	415	1326
310	929	392	1321
320	944	368	1312
330	959	342	1301
340	971	315	1286
350	981	287	1268
360	990	258	1248
370	997	229	1226
380	1003	199	1202
390	1007	168	1175
400	1010	138	1148
410	1012	107	1119
420	1013	77	1090
430	1012	47	1059
440	1011	17	1028
450	−1009	+ 13	+996
460	1005	41	964
470	1002	69	933
480	997	96	901

Table 2 (*continued*)

Mireds $(T°\text{к.})^{-1} \times 10^6$ $(C_2 = 14384\cdot8)$	Change in trichromatic coefficients if C_2 is altered from 14384·8 to 14320		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
500	−986	−149	+ 837
510	980	174	806
520	973	198	775
530	966	221	745
540	959	244	715
550	952	266	686
560	944	286	658
570	936	306	630
580	928	325	603
590	920	343	577
600	912	360	552
610	904	377	527
620	896	392	504
630	888	407	481
640	880	421	459
650	871	434	437
660	863	447	416

Table 3

Mireds $(T°\text{к.})^{-1} \times 10^6$ $(C_2 = 14384\cdot8)$	Change in trichromatic coefficients if C_2 is altered from 14384·8 to 14350		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
0	− 0	− 0	+ 0
10·	7	10	17
20	15	22	37
30	25	36	61
40	36	51	87
50	49	67	116
60	64	84	148
70	79	103	182
80	96	121	217
90	114	140	254
100	133	158	291
110	153	175	328
120	173	192	365
130	194	208	402
140	215	222	437

H. G. W. Harding

Table 3 (*continued*)

Mireds $(T°\text{к.})^{-1} \times 10^6$ $(C_2 = 14384\cdot8)$	Change in trichromatic coefficients if C_2 is altered from 14384·8 to 14350		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
150	−237	+234	+471
160	258	246	504
170	279	255	534
180	299	263	562
190	320	268	588
200	340	272	612
210	358	275	633
220	377	275	652
230	394	274	668
240	411	271	682
250	426	266	692
260	441	260	701
270	454	253	707
280	467	244	711
290	479	233	712
300	489	222	711
310	499	210	709
320	507	197	704
330	515	183	698
340	521	169	690
350	526	154	680
360	531	138	669
370	535	122	657
380	538	106	644
390	540	90	630
400	542	73	615
410	543	57	600
420	544	41	585
430	544	25	569
440	543	9	552
450	542	7	535
460	540	22	518
470	538	37	501
480	535	52	483
490	532	66	466
500	529	81	448
510	526	94	432
520	522	107	415
530	519	119	400
540	515	131	384
550	511	143	368
560	507	154	353
570	503	165	338
580	498	175	323

Table 3 (*continued*)

Mireds $(T°\text{K.})^{-1} \times 10^6$ $(C_2 = 14384 \cdot 8)$	Change in trichromatic coefficients if C_2 is altered from 14384·8 in 14350		
	$x \times 10^6$	$y \times 10^6$	$z \times 10^6$
600	−489	+194	+295
610	485	202	283
620	481	211	270
630	476	219	257
640	472	226	246
650	467	233	234
660	463	240	223

HYSTERESIS AND EDDY LOSSES IN SINGLE CRYSTALS OF AN ALLOY OF IRON AND SILICON

By A. J. C. WILSON,

Cavendish Laboratory, Cambridge:. now at University College, Cardiff

ABSTRACT. The total energy dissipated, in an alternating magnetic field, in single crystals of iron containing 2·1 % silicon has been measured calorimetrically. Results are given for the field applied in the three crystallographic directions [100], [110], [111]. The total losses are separated into the conventional " hysteresis " and " eddy " components by means of measurements at different frequencies. The " eddy " losses do not appear to be dependent on direction ; the " hysteresis " losses for [100] are about one-third of those for the other directions measured. A tentative explanation is suggested.

§1. INTRODUCTION

THE calorimetric method is convenient for measuring the rate of continuous heat production, even when the rate is small. A striking application was the measurement of the energy of beta-disintegration of Radium E (Ellis and Wooster, 1928). Other examples of its use are the measurement of losses in dielectrics at radio frequencies (MacGregor Morris and Grisedale, 1939) and the study of the breakdown of dielectrics under high voltage (Whitehead and Nethercot, 1935). Bates and his co-workers have applied a calorimetric method to the somewhat different problem of the energy changes taking place in various parts of a single magnetic cycle (Bates and Weston, 1941 ; Bates and Healey, 1943).

In an investigation of electrical sheet steels undertaken at the Cavendish Laboratory for the Electrical Research Association, it was necessary to devise

means for measuring energy losses in small samples of material ($\sim$5 grams).
The normal methods require considerably more material, so the possibilities
of using a calorimetric method were investigated. It has proved somewhat
slow for routine use, but results have been obtained on single crystals mag-
netized in three important crystallographic directions.

The principle of the method is very simple. A thermocouple is enclosed
in a pack of two or more sheets, and the increase in temperature, caused by
applying an alternating field for a known time is noted. The loss in ergs
per gram per cycle is then given by

$$L = \frac{C\theta}{Mft},\qquad\qquad \ldots\ldots (1)$$

where C is the heat capacity of the assembly in ergs per degree, θ is the tem-
perature increase, t is the time of application of the field in seconds, M is the
mass of the iron and f is the frequency. Two difficulties prevent the use of
equation (1) in this simple form; first, the observed temperature increase is
less than it should be because of heat losses during the time of application
of the field, and secondly, the heat capacity of the assembly is doubtful, because
of uncertainties in the heat capacity of Fe-Si alloys and lack of knowledge
as to how much of the thermocouple leads and the surrounding air is effectively
part of the assembly. It is therefore necessary to correct the observed tem-
perature rise, and to calibrate the assembly by means of a small heating coil
incorporated for the purpose. This calibration has the advantage that the
thermocouple acts only as a transfer instrument, and it is unnecessary to
know the factor for converting change in e.m.f. to change in temperature.

§2. APPARATUS, SPECIMEN AND MEASUREMENTS

The " calorimeter " was constructed so as to have as small a heat capacity
as possible; it consisted merely of six junctions of a copper-constantan thermo-
element, a 1-ohm constantan heating element, and a coil to measure the mag-
netic induction in the specimen. The junctions and the heating element
were placed side by side, making a flat plate about 1 cm. square, and held in
position with cotton thread and varnish. The coil was wound to a length
of about 0·4 cm. on a rectangular former about 0·25 by 2·8 cm., varnished to
prevent it from uncoiling, and then fixed to the thermocouple leads so that
the junctions and heater were supported in the middle of it. There were
thus two pockets, one on each side of the junctions and heater, into which
the specimen to be investigated could be put.

The specimen consisted of four single-crystal discs of Fe-Si, cut from a
single crystal about 7 cm. across. After the discs had been annealed at 1000°C.
for half an hour, their orientation was found by means of back-reflection Laue
photographs. It was found that [100] made an angle of 26°±2° with the
rolling direction, and that the plane of the sheet was within 6° of (011). The
lattice parameter of a portion of the large crystal was 2·8581 kx. (2·8638 A.),
corresponding to a silicon content of 2·1%. The average diameter of the
discs was 2·42$_0$ cm., the measured total thickness 0·15$_9$ cm., and the total
mass 5·32$_5$ g. The thickness calculated from the mass, density and diameter

would be about 0.15_1 cm. A pair of the discs, separated by a sheet of thin paper to prevent electrical contact, was placed in each of the pockets mentioned above, and secured by tying with fine thread. The whole assembly was suspended in a Dewar flask and placed in a solenoid.

The flux-measuring coil was connected directly to a rectifier voltmeter whose resistance was high compared with that of the coil. It is easily shown that for an ideal rectifier the maximum flux linking the coil is given by

$$N = \frac{10^8 E}{\sqrt{2} \pi f},\qquad\qquad \ldots\ldots(2)$$

where E is the voltmeter reading, and that therefore

$$4\pi I' = \frac{10^8 E}{\sqrt{2} \pi f n A},\qquad\qquad \ldots\ldots(3)$$

where I' is the maximum magnetic intensity uncorrected for the field, n is the number of turns in the coil and A is the area of cross-section of the iron. The value of $4\pi I'$ needs correction for the actual magnetic field; the correction was zero within the limits of error for $4\pi I$ less than about 10000, and amounted to about 300 at $4\pi I = 15000$. The apparent fields varied from 150 to 1000. The largest error in the corrected values of $4\pi I$ is uncertainty in what ought to be taken as the area of cross-section of the discs, since the measured thickness is about 5% greater than the calculated thickness. The measured thickness was used, so that the values of $4\pi I$ may all be low by 5%. In addition, there are individual errors due to the calibration of the voltmeter, uncertainty in the demagnetizing factor and the like. It is estimated that the values of $4\pi I$ are relatively correct within 3% and absolutely within 8%.

Before the assembly is heated, either by an alternating magnetic field or the heating coil incorporated, the temperature was nearly constant. During heating it rose almost linearly, and after heating fell fairly rapidly. The actual temperature rise was therefore somewhat less than that corresponding to the heat input, and a correction was made by extrapolating the initial and final portions of the temperature-time curve to the mid-point of the heating period. This was a simple straight-line extrapolation for the initial portion of the curve, but the fall was so steep in the final portion that analytical extrapolation had to be used. The fall was found to be exponential well within the limits of error, and a graphical method was devised for finding the constant of the curve rapidly. From observations of the temperature rise for a measured input to the heating element and the temperature rise produced by the magnetic field in a known time, the energy loss per magnetic cycle is easily calculated. The time of application of the field was usually 180 seconds. The largest errors in the losses so found would seem to be due to the extrapolation of the final portion of the time-temperature curve to the mid-point of the heating period. The percentage error produced depends somewhat on the temperature interval, increasing if it is made unduly small or unduly large. For most of the measurements it may be estimated at 3 to 4%. The error in the calibration measurements will of course affect all determinations of the energy loss in the same way, and so produce a systematic error, estimated at 2 to 3%. Other

sources of error in the calorimetric measurements are small compared with these two; it is estimated, therefore, that the energy losses are relatively correct in general within 5% and absolutely within 8%.

§3. RESULTS

Measurements of the total loss per cycle were made with the applied field in the plane of the discs and having as nearly as possible the directions [100], [110], [111], at values of $4\pi I$ from about 5000 to 15000, and at three different

Table 1. Smoothed losses at integral values of $4\pi I$
(Extrapolated values are enclosed in parentheses)

		Loss, ergs per gram per cycle						
$4\pi I$		3000	5000	7500	10000	12500	15000	17500
Direction	Frequency							
[100]	38	—	(41)	91	156	(240)	(340)	—
	50	17	46	103	181	280	406	586
	82	24	63	143	248	(380)	(540)	—
	121	—	81	180	316	496	787	(1150)
[110]	50	39	92	185	303	461	700	977
	81	—	(98)	210	359	578	916	(1480)
	145	57	145	302	510	809	1236	(1780)
[111]	50	31	87	197	358	537	741	(980)
	85	(34)	108	225	405	670	(1020)	—
	145	(48)	145	298	527	891	1390	(2040)

frequencies. From these measurements values of the total loss at integral values of $4\pi I$ were interpolated. It was found convenient to use graphs of log (total loss) plotted against log $(4\pi I)$, as such plots are nearly linear, whereas plots of total loss against $4\pi I$ are strongly curved. The interpolated values

Table 2. "Hysteresis" loss and "eddy" loss at 50$\sim$
(Extrapolated values are enclosed in parentheses)

		Ergs per gram per cycle						
	$4\pi I$	3000	5000	7500	10000	12500	15000	17500
	Direction							
Hysteresis loss	[100]	6	20	48	79	120	140	(190)
	[110]	29	68	120	200	280	400	(480)
	[111]	(22)	55	140	270	350	390	(420 ?)
Eddy loss per cycle at 50$\sim$	[100]	11	26	55	100	160	270	(400)
	[110]	10	24	62	110	180	300	(500)
	[111]	(9)	32	57	90	180	350	(560 ?)
	Mean	10	27	58	100	170	310	—

are given in table 1. A few values involving a slight extrapolation are also given in the table, but are distinguished by being enclosed in parentheses.

The total loss per cycle for a given value of $4\pi I$ is, as is usually found, nearly a linear function of frequency. Several typical plots are shown in figure 1.

The loss extrapolated to zero frequency is conventionally attributed to hysteresis, and the linear increase with frequency to eddy currents. The losses separated in this way are given in table 2. The relative errors are estimated to be not
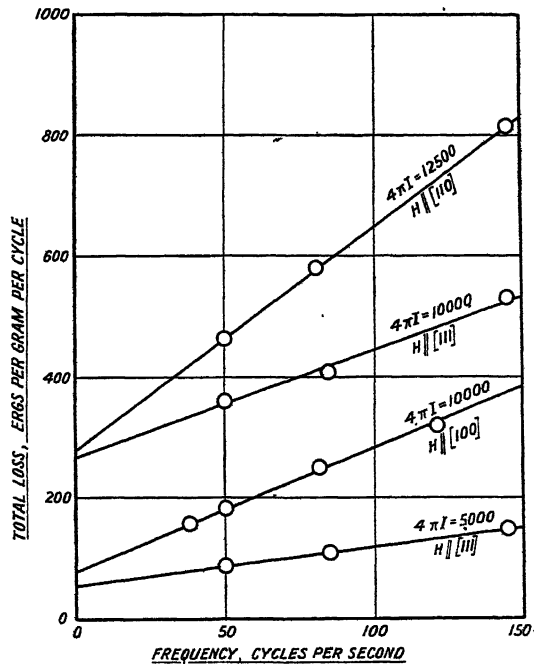

Figure 1. Extrapolation curves for hysteresis loss.

more than 10%. It will be noticed that the "hysteresis" component varies markedly with crystallographic direction, being almost three times as great for [110] or [111] as for [100], while the "eddy" component is nearly independent of direction, at any rate for $4\pi I$ less than 13000.

§4. DISCUSSION OF RESULTS

" Eddy " losses

It seems somewhat doubtful if the "eddy" component of the losses is due entirely to eddy currents. The heating due to this cause can be calculated with fair certainty; the theoretical formula for the loss per unit volume per second at low frequencies (flux density uniform throughout the sheet) is

$$W = \frac{\sigma b^2}{12 c^2} \overline{\left(\frac{\partial B}{\partial t}\right)^2}, \qquad \ldots\ldots(4)$$

where σ is the conductivity of the material, $\partial B/\partial t$ is time rate of change of the magnetic induction (the bar represents the time average over a complete cycle), b is the thickness of the sheet, and c is the speed of light. For a sinusoidal variation of B at $50\sim$ this becomes

$$W = \frac{82 b^2 B^2_{\text{max}}}{\rho\delta} \text{ ergs per gram per cycle,} \qquad \ldots\ldots(5)$$

where ρ is the resistivity of the sheet and δ its density, both in c.g.s. units. For the sheet used, $b = 0.040$ cm., $\rho \sim 35000$, $\delta \sim 7.7$, so the loss at $B_{max} = 10000$ should be about 49 erg/gm. cycle. The observed value is 100 erg/gm. cycle, which is almost exactly twice as big. A discrepancy of this order is usually found in such comparisons (Brailsford, 1943), and has not yet been satisfactorily explained. In some cases it is due, at least in part, to non-sinusoidal variation of B, but it may also be necessary to postulate some magnetic viscosity effect. In the present experiments B was very nearly sinusoidal, since discs have a very large demagnetizing factor and the field in the absence of the specimen was sinusoidal.*

The "eddy" losses are very similar for the three directions, as is shown in figure 2. Up to $4\pi I \sim 12500$ they are identical within the estimated limits of error ; at 15000 they differ by scarcely more. The separation of "eddy"
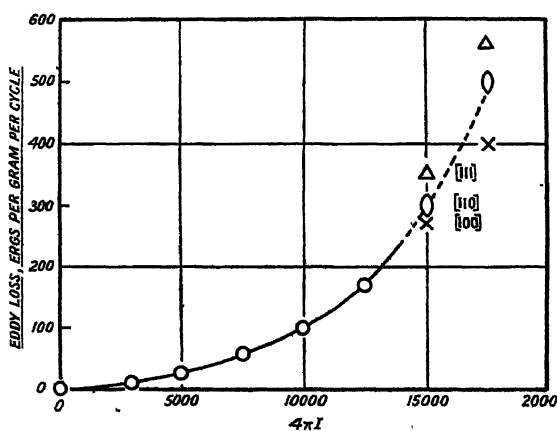
Figure 2. Eddy losses for the three directions at 50$\sim$.
Up to $4\pi I = 12500$ the losses are hardly separated on this scale.

and "hysteresis" losses at 17500 is little better than guess-work, but it confirms the impression given by the losses at 15000 that at high inductions the "eddy" losses increase in the order [100] < [110] < [111]. It seems doubtful if eddy losses in the strict sense would behave in this way; magneto-resistance effects should not amount to more than a few tenths per cent. This is perhaps confirmatory of the hypothesis that "eddy" losses include magnetic viscosity effects dependent on the frequency as well as eddy losses proper, but Richer (1944) has advanced a theory which would suggest that eddy currents ought to depend on the direction of magnetization.

" Hysteresis " losses

As has been seen, the losses in the [100] direction are much smaller than those in the [110] and [111] directions, in fact only about one-third as big. There is no great difference between the [110] and [111] directions ; [111] tends to be higher in the centre of the range, [110] at the ends. The following

* The effect of a large demagnetizing factor is to reduce the actual field to a very small value, in such a way as to make the magnetic induction in the specimen very nearly proportional to the value of the field in the absence of the specimen.

tentative discussion of the relation between the losses suggests itself. The argument is similar to that of Richer for eddy currents.

The first step in magnetization along any non-[100] direction is to increase the number of domains whose [100] magnetization makes a small angle with the field direction, and decrease the number of domains whose [100] magnetization makes a large angle with the field direction. Now, all [100] directions are symmetrically placed with respect to [111], and one may therefore suppose that all [100] directions behave similarly, and, indeed, as they would behave if the field were in a [100] direction, except that it will be the component of the field in the [100] direction that is important. An externally measurable value of I in a [111] direction will then correspond to an actual internal value of I in the [100] directions that is $\sqrt{3}$ times as big. The curve of hysteresis loss against $4\pi I$ might therefore be expected to coincide with the similar curve for [100] if the values of $4\pi I$ are multiplied by $\sqrt{3}$. The postulated mechanism
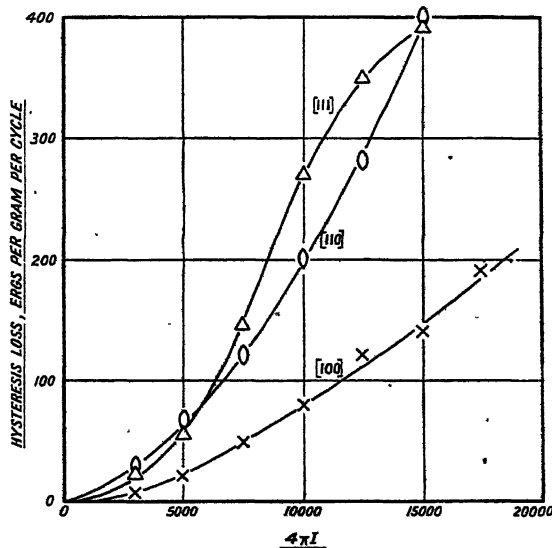
Figure 3. Hysteresis losses for the three directions.

for magnetization in the [111] direction is almost certainly that predominant for values of the externally-measured $4\pi I$ less than about $1/\sqrt{3}$ of the saturation value, so that the relation between the [111] and the [100] losses should hold over the entire range for which the latter have been measured.

The behaviour of the [110] direction is more difficult to predict. The [100] directions are not symmetrically placed with respect to [110] : two-thirds of them make angles of 45° (or 135°), one-third of them make angles of 90°. The assumption that immediately suggests itself is that the domains making angles of 90° with [110] make a 90° shift into one of the directions making angles of 45°. One might therefore expect that the curve of hysteresis loss againt $4\pi I$ would coincide with the similar curve for [100] if the values of $4\pi I$ are multiplied by $\sqrt{2}$. As in the case of [111], the relation should hold over the measured range of [100] losses.

In figure 3 the measured hysteresis losses are plotted against the externally

measured value of $4\pi I$ for the three directions, and in figure 4 they are re-plotted with the co-ordinates treated as the foregoing argument suggests : [100] plotted without change (crosses), [110] plotted with the polarizations multiplied by $\sqrt{2}$ (ellipses with long axes vertical), [111] plotted with the polarizations multiplied by $\sqrt{3}$ (triangles). The vertical lines are of a length corresponding to errors of 10%. It will be seen that up to $4\pi I = 15000$ the curves for [111] and [100] agree within this amount.

The agreement is quite impressive, and may perhaps be taken as some confirmation of the not quite rigorous theoretical argument. On the other hand [110] shows no agreement at all, and this probably indicates that the assumption made about the behaviour of the domains making angles of 90° with the [110] direction is erroneous. Qualitatively one may say that these domains are less effective in increasing the polarization than if they behaved
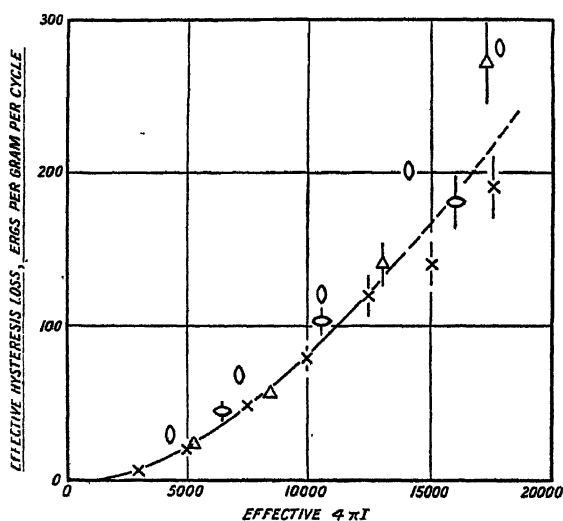
Figure 4.   Comparable hysteresis losses for the three directions.
For explanation of the points see text.

as assumed. By what is probably no more than a curious coincidence, the assumption of the opposite extreme (that these domains are entirely ineffective) gives good quantitative agreement. On this assumption one would have two-thirds of the iron magnetized to an intensity of $(3/2)\sqrt{2}$ times the externally measured intensity, and the curve of hysteresis loss against $4\pi I$ should coincide with that for [100] if the losses are multiplied by $3/2$ and the values of $4\pi I$ by $3/\sqrt{2}$. The [110] losses treated in this way are also plotted in the figure (ellipses with long axes horizontal), and it will be seen that they agree well with those for [100] and [111].

This explanation of the apparently anomalous behaviour of the [110] losses is confirmed by the tendency, postulated by Becker and Döring (1939), for the domains whose directions of magnetization lie nearly in the plane of the sheet to be favoured at the expense of those whose directions of magnetization make large angles with the plane of the sheet. In the present case domains

effective for the [100] measurements have their directions in the plane of the sheet, whereas those effective for the [110] measurements have their directions of magnetization at 45° to the plane of the sheet. This explanation might be regarded as qualitatively satisfactory, but the fraction one-third would be the result of coincidence. A little consideration shows that the [111] losses would not be affected if the greater part of, or even all, the domains had their direction of magnetization in the plane of the sheet. Richer (private communication), however, has suggested that the apparently anomalous behaviour of the [110] losses might be explained by a non-isotropic distribution of interstitial impurities set up in the process of rolling the sheet.

ACKNOWLEDGMENTS

REFERENCES

BATES, L. F. and HEALEY, D. R., 1943. *Proc. Phys. Soc.* 55, 188–202.
BATES, L. F. and WESTON, J. C., 1941. *Proc. Phys. Soc.* 53, 5–34.
BECKER, R. and DÖRING, W., 1939. *Ferromagnetismus*, p. 153.
BRAILSFORD, F., 1943. *J. Instn. Elect. Engrs.* 90, 307.
ELLIS, C. D. and WOOSTER, W. A., 1928. *Proc. Roy. Soc.* A, 117, 109.
MacGREGOR MORRIS, J. T. and GRISEDALE, G. L., 1939. *Phil. Mag.* 28, 34–63.
RICHER, G. C., 1944. *Report* T.C. 31 of Lysaght-Sankey Technical and Research Committee.
WHITEHEAD, S. and NETHERCOT, W., 1934. *Proc. Phys. Soc.* 47, 974.

COMPLETE COMPUTATION OF ELECTRON OPTICAL SYSTEMS

By H. MOTZ AND LAURA KLANFER,

Engineering Laboratory, Oxford

ABSTRACT. This paper shows how electron-optical systems can be completely calculated, without appeal to model experiments, by methods capable of great accuracy.

The computation method is a numerical one. It is explained with reference to the example of an electrostatic lens consisting of two cylinders with diameter ratio 1·5. The electric potential and the field are found by relaxation methods.

Electron paths for paraxial and non-paraxial electrons are computed by means of a step-by-step integration method which allows a close check on accuracy. The position of focal and cardinal points and the spherical aberration of the lens are found to be in fair agreement with experimental and semi-empirical determinations by other authors.

§1. INTRODUCTION

THE theoretical treatment of any electron-optical system proceeds in two steps : calculation of the electric or magnetic field of the focusing system and determination of the path of charged particles in these fields. In some cases analytic expressions for the fields are known, but in the majority of arrangements in practical use this is not the case. It is then usual to determine the field by experiment, either with the help of an electrolytic tank or by other means.

When analytic expressions for the field are known, the second step, path determination, can sometimes be carried out analytically. Again, in most cases this is not possible, and approximation methods have to be used. In a recent paper L. S. Goddard (1944) has given a summary of path-tracing methods. In that paper a step-by-step method of path computation is proposed which is used by Klemperer and Goddard (1944) for treatment of a magnetic lens. L. J. Comrie pointed out in the discussion that other step-by-step methods may be preferable. Goddard also indicates that the relaxation method may be used for the first step, the field computation.

We have been using both the relaxation method and a step-by-step method proposed by L. J. Comrie during the last few years for electron-optical calculations, and it will be shown in this paper that a complete numerical computation of electron-optical systems, without appeal to experiment, is practicable.

An electrostatic lens will be taken as an example, and the agreement of the result of the determination of focal length with the results of other authors will be discussed.

The methods used are, *mutatis mutandis*, equally applicable to axially symmetrical magnetic focusing systems. Even saturation effects in ironclad coils can be taken into account. A discussion of magnetic field computation will

shortly appear in the *Journal of the Institution of Electrical Engineers*, where the application of the relaxation method to potential problems will be fully explained.

§2. THE FINITE-DIFFERENCE METHOD OF FIELD CALCULATION

For the purpose of the field determination, Laplace's equation,

$$\frac{\partial^2\phi}{\partial r^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} + \frac{\partial^2\phi}{\partial z^2} = 0, \qquad \ldots\ldots(1)$$

for the potential ϕ has to be solved with certain boundary conditions. For this equation a system of difference equations for the values of ϕ at points of a net of small but finite mesh length is substituted. The choice of the net depends on the shape of the boundary. In the present paper a study is made of an electrostatic lens, consisting of two semi-infinite cylinders placed end to end and separated by a finite gap. The cylinders as shown in figure 1 have diameters D_1 and D_2, with $D_2/D_1 = 1\cdot5$, and their potentials are 200 v. and 1200 v. respectively. This choice fixes the boundary conditions. The voltage ratio is thus $V_2/V_1 = 6$, but when a solution is found for this ratio, a solution for any other voltage ratio can be obtained quite simply (see Maloff and Epstein, 1938, p. 102). Three quadratic nets of successively finer mesh length are shown in figure 3. Consider at first the coarsest net, of mesh-length a. At any point P of this net the potential ϕ can be expressed in terms of the potentials ϕ_Q, ϕ_R, ϕ_S, ϕ_T of the four neighbouring mesh points, Q, R, S, T, figure 2. It can be seen, with the help of an expansion in Taylor series of the potential at these points, that

$$\phi_Q\left(1 - \frac{1}{2\rho_P}\right) + \phi_S\left(1 + \frac{1}{2\rho_P}\right) + \phi_R + \phi_T - 4\phi_P = 0, \qquad \ldots\ldots(2)$$

where $\rho = r/a$ and $\rho_P = r_P/a$, so that r_P is the radial distance of point P from the axis. If there are n mesh points, excluding points on the boundary where $\phi = \text{const.}$, n simultaneous equations have to be satisfied. Owing to the symmetry of the figure, the equations on both sides of the axis are the same, so that only one-half of figure 1 is used for the computation.

A solution of these equations can be found by various methods, e.g. the Gauss-Doolittle elimination method or R. V. Southwell's relaxation method (Christopherson and Southwell, 1938). We have used the latter method, which we prefer because it is self-checking in so far as mistakes made at some stage of the calculation are corrected later with little loss of time and effort. At any stage of the process, the residuals

$$\phi_Q\left(1 - \frac{1}{2\rho_P}\right) + \phi_S\left(1 + \frac{1}{2\rho_P}\right) + \phi_R + \phi_T - 4\rho_P = \delta_P \qquad \ldots\ldots(3)$$

are found and reduced in successive steps until they are negligible. A further description of the process is given in a forthcoming paper by Motz and Worthy in the *Journal of the Institution of Electrical Engineers*.

At the end a further test should be applied. The system of finite-difference equations (2) is accurate only to the second order in a, and we are only justified in using it when ϕ does not change rapidly in the neighbourhood of any

H. Motz and Laura Klanfer

mesh-point. It is possible to locate points or regions where second-order approximation is not good enough by forming the higher differences of ϕ in rows and
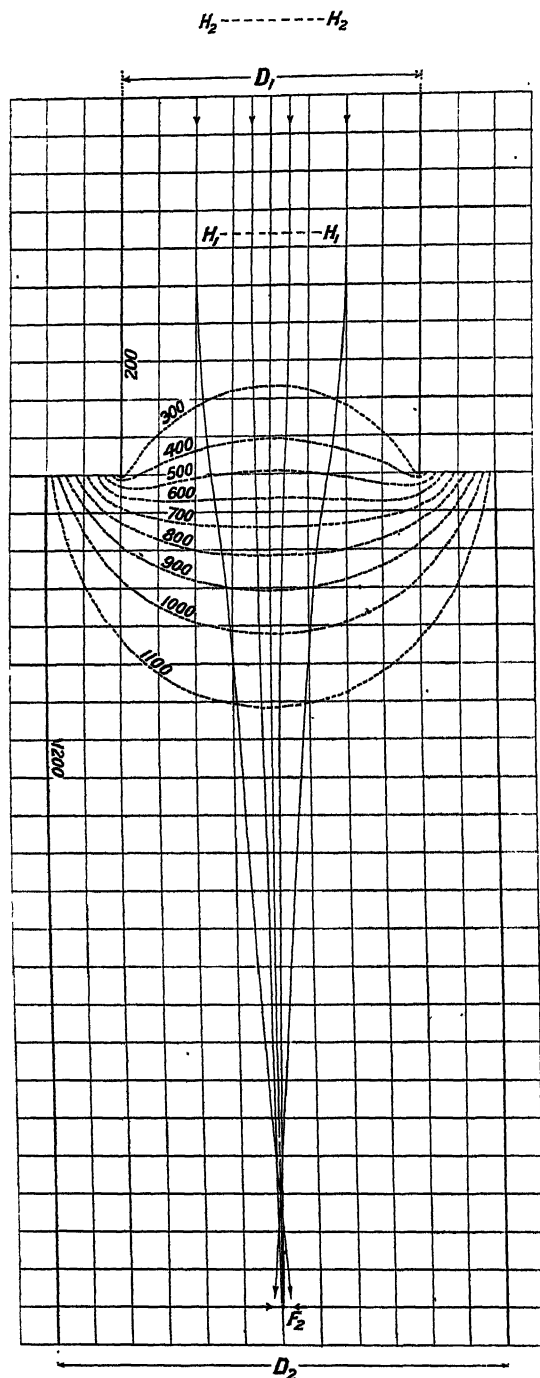

Figure 1.

columns. Failure of the second-order approximation is disclosed by erratic behaviour of the higher differences. Consider, for example, the effect of a change

of ϕ at one mesh point by ± 1 on the higher differences. If the fourth differences fluctuate by an amount larger than the one which would be produced by errors of ± 1 in the last significant figure of the result, it must be concluded that this

375	371	364	353	336	313	283	245	200		
396	394	387	375	357	332	298	253			
397	395	388	377	360	336	302	256			
397	395	383	377	360	336	304	258			
421	419	412	401	382	355	317	265	200		
447	445	433	427	409	383	344	285			
449	447	441	431	415	390	352	292			
449	447	441	431	416	392	356	297			
477	475	469	459	442	418	376	302			
505	503	498	489	475	453	418	352	200		1200
508	507	503	495	483	464	433	378			
508	507	503	495	485	471	444	405			
543	542	538	532	523	510	491	455	465		
568	567	564	560	553	543	531	520	532		1200
574	573	571	567	562	555	546	539	564		
576	575	574	569	567	564	558	549	595		
613	613	612	611	611	611	618	627	667		
635	635	635	635	636	638	646	666	713		1200
642	642	642	643	645	649	659	680	728		
646	646	645	648	652	658	671	690	738		
682	683	685	688	694	701	717	747	798		
703	704	706	710	717	728	746	778	830		1200
710	711	713	717	725	737	756	788	839		
715	716	718	721	731	743	763	794	843		
750	751	755	761	771	786	808	840	886		
769	770	774	781	791	807	830	864	912		1200
776	777	780	787	798	814	838	871	918		
781	782	784	790	803	819	844	875	919		
813	814	819	827	839	856	879	912	954		
830	832	837	845	857	874	898	931	973		1200
837	838	842	850	863	880	904	936	977		
842	842	844	852	867	884	908	938	977		
871	872	877	885	898	915	938	968	1004		
886	888	893	901	914	931	954	983	1018		1200
892	893	898	906	919	936	958	987	1021		
896	896	901	909	922	939	959	987	1021		
966		972		991		1025		1074		
978		984		1003		1036		1082		1200
982		988		1006		1038		1084		
985		991		1007		1038		1085		

Figure 2.

figure is uncertain by more than ± 1. Difference formulae, of higher than second-order accuracy, may then be used, or the net may be subdivided.

It is clear that the cylinder edges, figure 1, are singularities for ϕ, and it can be shown that near them ϕ varies as the square root of the distance from the

edge. Hence in the neighbourhood of sharp edges, the higher-order residuals cannot be neglected. Methods for dealing with this singularity have been evolved which are based on the behaviour of the potential near the singularity, but in this paper the simplest expedient has been used. It consists in the use of two successive finer nets as shown in figure 1. When the net is subdivided in a particular region, the computation is carried on in alternative steps (i) and (ii).

(i) Values of the potential at fine mesh points on the boundary between the fine and coarse net region are found by interpolation. ϕ-values on this boundary are then kept constant and, with these values, ϕ is found at points of the fine-mesh region by the relaxation process. Some of those points belong also to the coarse net.

(ii) ϕ-values at these latter points are next kept constant and new ϕ-values in the coarse net are found. The fine-net values are redetermined according to (i), etc., until the process converges. The result of the computation with the coarse, medium and finest net are retained. Figure 2 shows the middle region of figure 1 with the medium net. The three successive ϕ-values are recorded and the fourth figure is found by parabolic extrapolation to mesh-length zero. At points not too near the edges the convergence is good. The extrapolation has only been carried out in a region where electron paths have been computed. In any practical focusing system the wall thickness of cylinders is finite. The mesh length of the finest net should be of the order of the radius of curvature of the edge. It will be seen later that the focal lengths of a lens with a wall thickness given by our finest net differ insignificantly from those of an ideal lens having cylinder walls of zero thickness.

§3. TREATMENT OF THE REGION OUTSIDE THE CYLINDERS

In any practical system, electrostatic shields will be found outside the focusing electrodes. We shall later compare the results of our path computation with those given by Maloff and Epstein (1938) and by Spangenberg and Field (1943). Neither of these papers gives information about shielding, and we have, therefore, not included a shield. To treat the infinite space outside, we have carried out an inversion with respect to a cylinder of radius $r_0 = 9a$. A new variable, $s = r_0/r$, is introduced.

Equation (1) becomes

$$\frac{\partial^2 \phi}{\partial z^2} + \frac{s^4}{r_0^2} \frac{\partial^2 \phi}{\partial s^2} + \frac{s^3}{r_0^2} \frac{\partial \phi}{\partial s} = 0. \qquad \ldots\ldots(4)$$

The mesh length b of the s-net outside the inversion cylinder was taken as $b = \frac{1}{4}$ (s is non-dimensional).

Corresponding values are then :

$s =$	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$r =$	∞	$36a$	$18a$	$12a$	$9a$

The cylinders are long, but finite, compared with D_1 and D_2. At large distances, the field due to the cylinders will behave as a dipole field. Thus at $r = \infty$, $\phi = 0$. The effect of the end charges of the cylinders is neglected, but it is clear that these
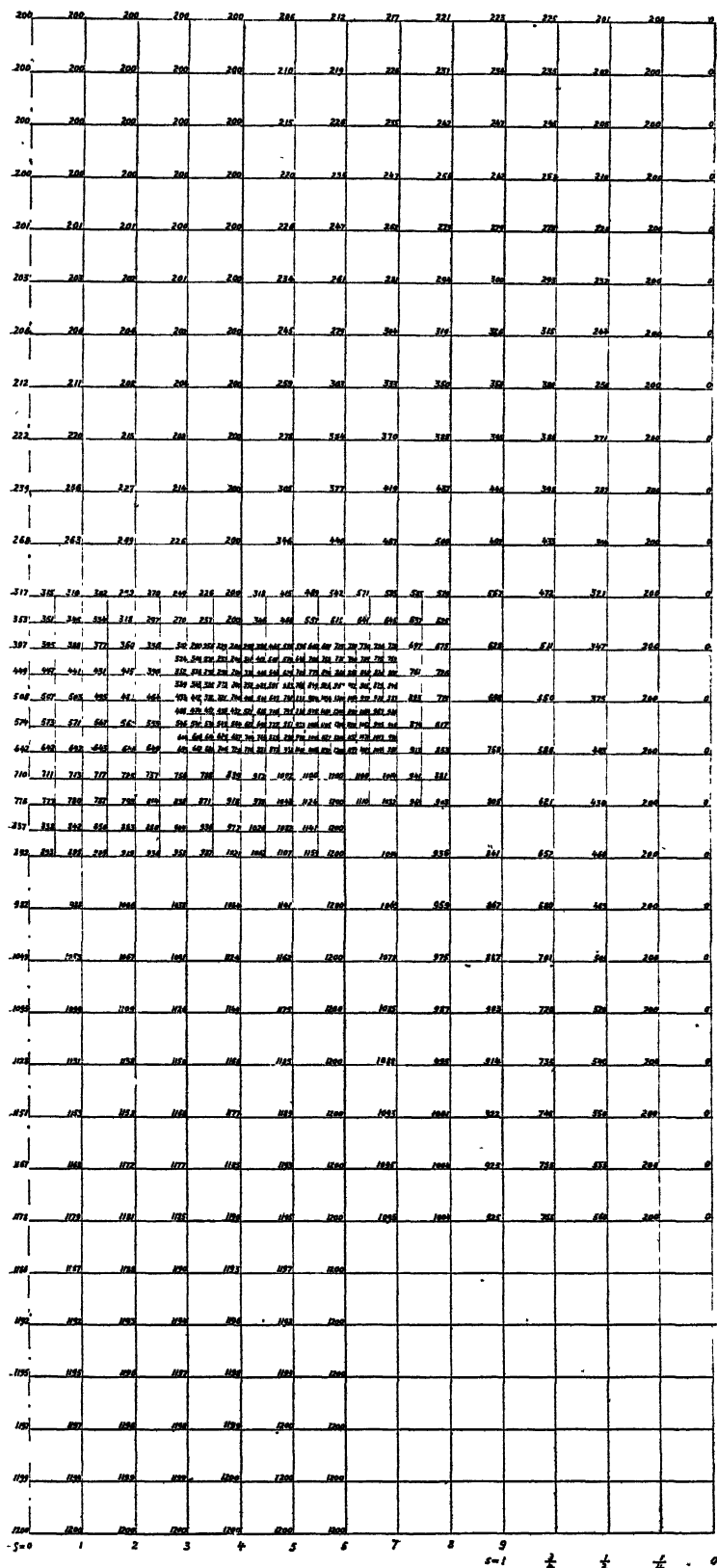
Figure 3.

3—2

effects must cancel at radial distances small compared with the length of the cylinders where we wish the potential to be accurately determined.*

The finite-difference equations corresponding to (4) become

$$\left(\frac{16}{81}\rho_P{}^4 - \frac{2}{81}\rho_P{}^3\right)\phi_Q + \left(\frac{16}{81}\rho_P{}^4 + \frac{2}{81}\rho_P{}^3\right)\phi_S + \phi_R + \phi_T - \left(\frac{32}{81}\rho_P{}^4 + 2\right)\phi_P = 0.$$
$$\cdots\cdots(5)$$

There is a slight ambiguity on the line $r = 9a$, where (2) or (5) may be used. Both (2) and (5) are difference equivalents of (1), and we have used the mean of the coefficients of these equations. The result of the potential calculations is shown in figure 3.

§4. DETERMINATION OF THE PATH

Two paraxial electron paths have been computed and one non-paraxial one. In figure 1 one paraxial path is seen entering the smaller cylinder at $r = D_1/16$. Another path entering the larger cylinder in a direction parallel to the axis at the same radial distance has also been computed. These two paths determine focal length and principal planes. The non-paraxial path gives a measure of the spherical aberration in the forward direction accelerating lens.

The equation for the path of a negatively charged particle is

$$\frac{d^2r}{dz^2} + \frac{\partial V}{\partial z}\frac{\left[1 + \left(\frac{dr}{dz}\right)^2\right]\frac{dr}{dz}}{2V} - \frac{1}{2V}\frac{\partial V}{\partial r}\left[1 + \left(\frac{dr}{dz}\right)^2\right] = 0. \quad\cdots\cdots(6)$$

The method of path computation is based on a process suggested by L. J. Comrie in *Interpolation and Allied Tables* (1936).

Define

$$F = 4\frac{d^2\rho}{d\zeta^2} \quad f = 2\frac{d\rho}{d\zeta},$$

where $\rho = r/a$, $\zeta = z/a$, and use steps

$$\rho_{n+1} - \rho_n = z_{n+1} - z_n = a.$$

Equation (6) can be written

$$F = -\frac{\partial V}{\partial z}\frac{\left(1 + \frac{f^2}{4}\right)}{V}f + 2\frac{\partial V}{\partial r}\frac{\left(1 + \frac{f^2}{4}\right)}{V}. \quad\cdots\cdots(7)$$

The procedure consists in computing entries of the following table :—

<div align="center">Table 1</div>

z/a	ρ	d'	d''	d'''	f	F	Δ'	Δ''
0	ρ_0		d_0''		f_0	F_0		Δ_0''
		$d_{1/2}'$		$d_{1/2}'''$			$\Delta_{1/2}'$	
1	ρ_1		d_1''		f_1	F_1		Δ_1''
		$d_{3/2}'$		$d_{3/2}'''$			$\Delta_{3/2}'$	
2	ρ_2		d_2''		f_2	F_2		Δ_2''
		$d_{5/2}'$		$d_{5/2}'''$			$\Delta_{5/2}'$	
3	ρ_3		d_3''		f_3	F_3		
.	.		.					.
.				.				

---

* Alternatively one might interpret the field as due to infinite cylinders and a zero potential hield at infinity.

where d', d'', d''' are successive central differences of ρ, and Δ', Δ'', Δ''', ... are successive central differences of F.

Assume ρ_0, ρ_1, ρ_2 known; in our example

$$\rho_0 = \rho_1 = \rho_2 = D_1/16.$$

Calculate f_0, f_1, f_2 from, for example,

$$f_1 = \left(2d_1'\right) - \frac{1}{6}\left(2d_1'''\right) + \frac{1}{30}\left(2d_1^{(V)}\right) - \dots \quad \dots \dots (8)$$

and F_0, F_1, F_2 from equation (7) and difference. Δ_1'' is already found in this way, but not Δ_2''.

$$\frac{1}{V}\frac{\partial \phi}{\partial z}, \quad \frac{1}{V}\frac{\partial \phi}{\partial r} \qquad \dots \dots (9)$$

are taken from the field plot, figure 3. (Tables of these values are prepared beforehand so that appropriate values of (9) can be interpolated for each value of ρ. For interpolation we use a four-point formula (Comrie, 1936, p. 685)).

From now on the following steps are carried out:

(i) Estimate Δ_2'', e.g. take $\Delta_2'' = \Delta_1''$ and calculate d_2'' from

$$d_2'' = \frac{1}{4}\left(F_2 + \frac{1}{12}\Delta_2'' - \frac{1}{240}\Delta_2^{(iv)} + \dots\right). \qquad \dots \dots (10)$$

(ii) Form ρ_3 by building up from d_2''.

(iii) Verify by calculating f_1 from (8) and, if necessary, make a correction.

(iv) Now extrapolate $f_3 = f_1 + F_2 + \frac{1}{6}\Delta_2'' - \frac{1}{180}\Delta_2^{(iv)} + \dots \quad \dots \dots (11)$

(v) Compute F_3 and difference;

(vi) Verify that correct value of d_2'' has been used in (ii); correct if necessary.

The steps of z used in the path computation are conveniently chosen so that $\frac{1}{48}$ of Δ_2'' is not larger than a few units of the last decimal figure of F which is retained during computation. It should be assumed that one significant figure is lost in about ten steps of the integration. It is therefore necessary; if absolute certainty of the result is desired, to retain during the computation more decimal figures than the final accuracy claimed for the path warrants.

If, in our table 2, the last figure of F is uncertain by ± 1, then the r might, in any step, acquire an error of one-third of the last figure retained. There are 30 steps; the errors accumulate in arithmetic progression so that, if all the errors are in the same direction, the final result may be in error by 0·150 and the position of the focal points in error by two mesh lengths (cf. the end of the path given in table 2). Now it is very unlikely that all the errors are in fact in the same direction, and we estimate that the error is less than one mesh length.

The process outlined above leads to table 2 for the paraxial path shown in figure 1. We have found it convenient to use the steps of the relaxation net for the path computation. Thus the first eight steps are carried out in the coarsest net. To start the path in the middle region medium net, two values of ρ are interpolated at mid-points of the last two steps. The last part, again in the coarse net, is started by selecting the last three values of ρ belonging to the coarse net in the

middle region. At the end the path is a straight line and the first differences are 0·147. The point of intersection is found by interpolation. The paraxial path was first determined in the coarse net throughout, and F_2, the distance of the focal

Table 2

ζ	ρ	d'	d''	d'''	f	F	$\varDelta'$	$\varDelta''$
−10	0·500		0·000		0·000	0·000		
		0·000		0·000			0·000	
− 9	0·500		−0·001		−0·001	0·000		−0·002
		−0·001		0·000			−0·002	
− 8	0·499		0·000		−0·002	−0·002		−0·001
		−0·001		−0·001			−0·003	
− 7	0·498		−0·001		−0·003	−0·005		−0·002
		−0·002		0·000			−0·005	
− 6	0·496		−0·002		−0·006	−0·010		0·000
		−0·004		−0·001			−0·005	
− 5	0·492		−0·004		−0·012	−0·015		−0·003
		−0·008		−0·002			−0·008	
− 4	0·484		−0·006		−0·021	−0·023		−0·001
		−0·014		−0·001			−0·006	
− 3	0·470		−0·007		−0·035	−0·029		+0·004
		−0·021		−0·001			−0·002	
− 2	0·449		−0·006		−0·048	−0·031		+0·009
		−0·027		0·000			+0·007	
− 1	0·422		−0·006		−0·061	−0·024		+0·015
		−0·033		+0·006			+0·022	
0	0·389		0·000		−0·067	+0·002		−0·010
		−0·033		+0·004			+0·012	
+ 1	0·356		+0·004		−0·063	+0·014		−0·012
		−0·029		−0·001			0·000	
+ 2	0·327		+0·003		−0·055	+0·014		0·000
		−0·026		+0·001			0·000	
+ 3	0·301		+0·004		−0·048	+0·014		+0·004
		−0·022		−0·002			+0·004	
+ 4	0·279		+0·002		−0·042	+0·009		−0·002
		−0·020		0·000			−0·002	
+ 5	0·259		+0·002		−0·038	+0·007		+0·001
		−0·018		−0·001			−0·003	
+ 6	0·241		+0·001		−0·035	+0·004		+0·002
		−0·017		0·000			−0·001	
+ 7	0·224		+0·001		−0·034	+0·003		0·000
		−0·016		0·000			−0·001	
+ 8	0·208		+0·0005		−0·032	+0·002		0·000
		−0·0155		0·000			−0·001	
+ 9	0·1925		+0·0003		−0·031	+0·0012		+0·001
		−0·0152		0·000			0·000	
+10	0·1773		+0·0002		−0·030	+0·0007		0·000
		−0·0150		0·000			0·000	
+11	0·1623		+0·0001		−0·030	+0·0005		0·000
		−0·0149		0·000			0·000	
+12	0·1474		+0·0001		−0·030	+0·0002		0·000
		−0·0148		0·000			0·000	
+13	0·1326		0·0000		−0·030	+0·0002		0·000
		−0·0148		0·000			0·000	
+14	0·1178		−0·0001		−0·030	+0·0002		0·000
		−0·0147 etc.						

Table 2—*continued*

Medium region computed in half-intervals

2ζ	2ρ	d'	d''	d'''	f	F	Δ'	Δ''
− 7	0·956		−0·000		−0·028	−0·013	interpolated	
		−0·016		−0·000			−0·002	
− 6	0·940		−0·004		−0·036	−0·015		+0·001
		−0·020		+0·002			−0·001	
− 5	0·920		−0·002		−0·042	−0·016	interpolated	+0·001
		−0·022		−0·002			0·000	
− 4	0·898		−0·004		−0·048	−0·016		+0·003
		−0·026		+0·001			+0·003	
− 3	0·872		−0·003		−0·055	−0·013		−0·002
		−0·029		0·000			+0·001	
− 2	0·843		−0·003		−0·061	−0·012		+0·005
		−0·032		+0·001			+0·006	
− 1	0·811		−0·002		−0·067	−0·006		+0·001
		−0·034		+0·002			+0·007	
0	0·777		0·000		−0·069	+0·001		−0·002
		−0·034		+0·002			+0·005	
+ 1	0·743		+0·002		−0·066	+0·006		−0·004
		−0·032		0·000			+0·001	
+ 2	0·711		+0·002		−0·062	+0·007	end path start	0·000
		−0·030		0·000			+0·001	
+ 3	0·681		+0·002		−0·058	+0·008		−0·002
		−0·028		0·000			−0·001	
+ 4	0·653		+0·002		−0·054	+0·007	end path	0·000
		−0·026		−0·001			−0·002	
+ 5	0·627		+0·001		−0·051	+0·005		
		−0·025						
+ 6	0·602				−0·049		end path	

point from the cylinder junction, was $F_2 = +15$. Then it was redetermined with the more exact figures of the medium net, and finally again with the best extrapolated figures; the last two results for F_2 were the same, although the paths were slightly different, and we therefore considered our result as the final answer. The distances of the focal planes F_1, F_2 and the cardinal planes H_1, H_2, from the plane through the cylinder junction, are given in table 3.

§5. COMPARISON WITH THE RESULTS OF OTHER AUTHORS

Maloff and Epstein (1938) on p. 105 give data for a lens similar to the one computed in this paper with $D_2/D_1 = 1·5$, $V_2/V_1 = 6$.

There is no precise information in the book concerning the method by which these results were obtained. In particular, no details are given about position and potential of shielding electrodes outside the lens either in the electrolytic trough experiments determining the field or in the experimental verification of focal length, etc., given on p. 114. It is, however, likely that one grounded shield was used. The wall thickness of the cylinders is not stated. All these data are also omitted by Spangenberg and Field (1942), p. 195, who give a chart for a similar lens with the same parameters. It must be concluded from the nature of the experimental arrangement that two shields at different potentials were used. The results taken from these papers are summarized in table 3.

It is difficult to say how much of the disagreement is due to the difference in shielding and wall thickness in the three cases. It will be seen that the distances between focal points agree wthin one mesh length, and this is within the accuracy

Table 3

	F_1	F_2	H_1	H_2
Maloff and Epstein	$-21 \cdot 6$	20	$-8 \cdot 8$	$-13 \cdot 2$
Spangenberg and Field	$-21 \cdot 68$	$18 \cdot 11$		
Our result : Paraxial	$-18 \cdot 56$	22	$-6 \cdot 37$	$-12 \cdot 01$
Non-paraxial		$19 \cdot 25$		

of our calculation. In our result the whole system of principal planes and focal planes seems to be shifted bodily by about three mesh lengths towards the high-potential electrode.
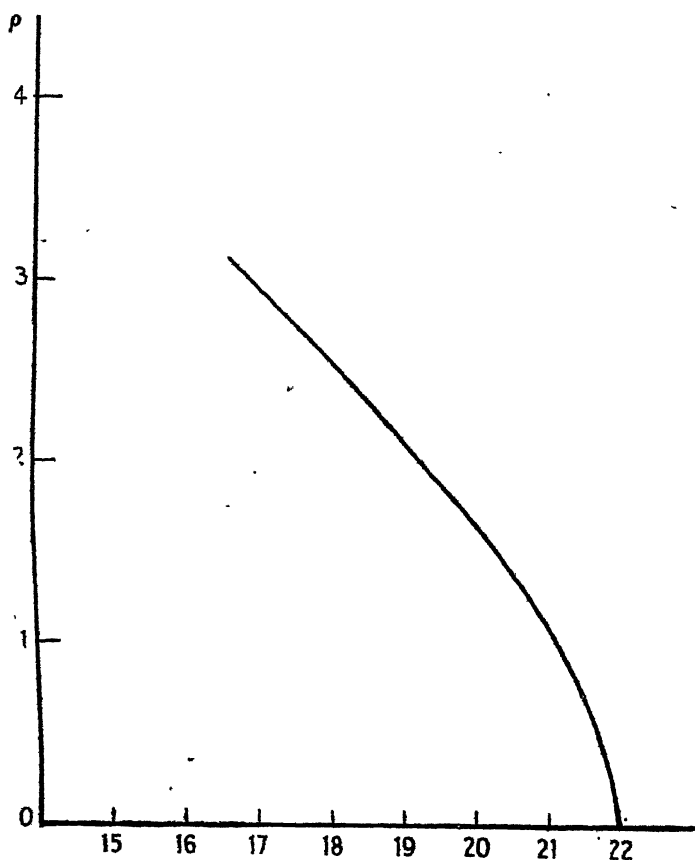
Figure 4.

Spangenberg and Field (1942) give on p. 312 a graph for the reduction of length for a non-paraxial path in an accelerating lens. They state that this reduction curve is almost universal if percentage reductions are considered. We have computed a non-paraxial path (figure 1 and table 3). The reduction

in focal length is 87·5%, whereas the reduction read off the universal curve mentioned above is 87%.

In order to obtain the full spherical-aberration curve, several other non-paraxial paths were computed, and the curve obtained is given in figure 4.

We conclude that our path determinations are internally consistent and that the differences between our results and those of the authors quoted are mostly due to errors in the field computation and the uncertainty of the geometrical data. We could, of course, increase the accuracy of our field determination. Such a further effort is not likely to effect our results materially and does not seem justified in view of the experimental errors.

REFERENCES

COMRIE, L. J., 1936. *Interpolation and Allied Tables* (London : H.M. Stationery Office).
GODDARD, L. S., 1944. *Proc. Phys. Soc.* 56, 342.
GODDARD, L. S. and KLEMPERER, O., 1944. *Proc. Phys. Soc.* 56, 348.
MALOFF, I. G. and EPSTEIN, D. W., 1938. *Electron Optics in Television* (New York and London : McGraw-Hill Publishing Co. Inc.).
MOTZ, H. and WORTHY, W. D. " Calculation of the Magnetic Field in Dynamo-electric Machines." (To appear shortly in *J. Instn. Elect. Engrs.*)
SPANGENBERG, K. and FIELD, LESTER M., 1942. *Electrical Communications*, 20, 305 ; *ibid.* 21, 194.

ADDITIONAL BIBLIOGRAPHY

CHRISTOPHERSON, D. G. and SOUTHWELL, R. V., 1938. *Proc. Roy. Soc.* A, 168, 317.
FOX, L., 1944. *Q. J. Appl. Math.* 2, 251.
SOUTHWELL, R. V. and VAISEY, G., 1943. *Proc. Roy. Soc.* A, 132, 129.

A MODIFIED HELMHOLTZ LINE-ELEMENT IN BRIGHTNESS-COLOUR SPACE

BY W. S. STILES,

Teddington

ABSTRACT. Helmholtz, Schrödinger and others have considered the relation between the trichromatic coordinates, x, y, z and $x+\delta x$ $y+\delta y$, $z+\delta z$, of two juxtaposed light patches which can just be discriminated by the eye. The Helmholtz expression,

$$\left(\frac{\delta x}{a+x}\right)^2 + \left(\frac{\delta y}{b+y}\right)^2 + \left(\frac{\delta z}{c+z}\right)^2 = \text{const.} = \delta s^2,$$

is the most readily interpretable in terms of the trichromatic theory, but to explain the observed variation of the hue limen through the spectrum it demands double-peak fundamental response curves which are at variance with other evidence. It also leads to a step-by-step visibility curve of wrong shape. The difficulties mentioned are removed by introducing different constant factors in the three terms. Such a modification is indicated independently by recent measurements of the liminal brightness increment, from which appropriate values of the factors have been derived. The factor in each term is simply related to the limiting Fechner fraction of the corresponding trichromatic mechanism, i.e., the Fechner fraction $\Delta B/B$ which would be observed at sufficiently high

brightnesses B if visual discrimination depended on that mechanism only. The modified expression for δs^2 is applied to the calculation of hue limens, the step-by-step visibility curve, the Fechner fraction curves, and the general colour limen at different points of the colour triangle. Certain main features of the experimental results are correctly reproduced, but discrepancies with Wright and MacAdam's measurements of general colour limens may indicate that the modified element ignores some factor which is operative in their experiments.

§1. INTRODUCTION

THE examination of the properties of the proposed line-element will be preceded by a brief comment on earlier work. It was Helmholtz (1891) who first attempted to express in mathematical form the condition for two juxtaposed light patches to be on the limit of discrimination by the eye. He suggested that any two just-distinguishable patches would have trichromatic co-ordinates x, y, z and $x+\delta x$, $y+\delta y$, $z+\delta z$ respectively, satisfying a relation of the form,

$$3\delta s^2 \equiv \left(\frac{\delta x}{a+x}\right)^2 + \left(\frac{\delta y}{b+y}\right)^2 + \left(\frac{\delta z}{c+z}\right)^2 = 3F^2, \qquad \ldots\ldots (1)*$$

where a, b, c and F are constants. The trichromatic co-ordinates here are the actual quantities of a particular set of primaries required to match the light patch—not the actual quantities divided by their sum (the unit co-ordinates). It is customary to regard the quantity δs defined in (1) as the length of the line-element in a non-Euclidean space of co-ordinates x, y, z. In this so-called brightness-colour space, the points corresponding to any pair of just-distinguishable light patches are the same elementary distance F apart. If a change is made to a different set of primaries and the relation (1) is expressed in terms of the new co-ordinates of the light patches, the simple form of the relation is lost and it is no longer the case that each term depends on one co-ordinate only. Thus, if the original form is valid at all, it is valid for only one set of the infinitely many sets of primaries which would serve equally well for the statement of the results of colour-matching measurements. The Helmholtz line-element implies, in fact, a particular set of primaries which, on the usual physiological interpretation of the trichromatic theory, we should expect to be the fundamental set such that each primary stimulates one, and only one, of the cone mechanisms. The quantities of the fundamental primaries required to match a monochromatic patch of unit energy intensity define three functions of wave-length—the spectral *sensitivity curves* or *fundamental response curves* of the three mechanisms.

Helmholtz took König's measurements of the trichromatic co-ordinates of the spectrum colours and sought to express them in terms of a set of primaries for which the relation (1) would be valid. The test was that the values of the hue limen through the spectrum as derived from (1) should agree with the measurements of this quantity made by König and Dieterici (1884). Helmholtz actually used the simpler form,

$$3\delta s^2 \equiv \left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2 + \left(\frac{\delta z}{z}\right)^2 = 3F^2, \qquad \ldots\ldots (2)$$

to which (1) reduces at sufficiently high intensities when a, b, c can be neglected compared with x, y, z respectively. He was able to find a set of primaries which

---

* The factor 3 is introduced so that F has a simple physical meaning, as explained below. Relation (1) is sometimes referred to as *the threshold condition*.

satisfied this test fairly well. But the corresponding fundamental response curves (reproduced as figure 8 in Peddie's *Colour Vision*) each have two pronounced maxima and are quite unlike the *Grundempfindungen* proposed by König from a consideration of the properties of colour-blinds. Moreover, as Schrödinger (1920) pointed out, they lead to an impossible double-peak form for the step-by-step visibility curve (the curve showing the relative energies of neighbouring spectrum colours which appear equally bright).

Schrödinger raised a further objection to the Helmholtz line-element. Although the step-by-step visibility curve can be derived from the line-element and its associated fundamental response curves, some additional assumption is necessary to determine when the eye will judge as equally bright two juxtaposed patches which differ widely in colour. Schrödinger assumed that two patches would appear equally bright if any change in the intensity of one of them would increase the minimum number of just-perceptible steps needed to pass from one to the other. The minumum number for any two light patches A and B is proportional to the integral

$$\int_A^B ds$$

taken along the geodesic in b.c. space. Thus the assumption is that A will match B when the intensity of B is adjusted to make this integral a minimum. While mathematically attractive, this definition of equality of brightness is not very plausible from a physiological standpoint. Applied to the Helmholtz element, it makes brightness a non-additive property of lights, in conflict with Abney's law. Schrödinger proposed an alternative line-element,

$$\delta s^2 \equiv \frac{1}{\alpha x + \beta y + \gamma z} \left[\frac{\alpha \delta x^2}{x} + \frac{\beta \delta y^2}{y} + \frac{\gamma \delta z^2}{z} \right] = F^2, \quad \ldots \ldots (3)$$

which, on his assumption, makes brightness proportional to $\alpha x + \beta y + \gamma z$ and, therefore, strictly additive. An adequate criticism of this line-element will not be attempted, but it appears to the writer inconsistent with the threshold measurements considered below.

Since Schrödinger's theoretical work, several investigators have determined by experiment sets of pairs of just-distinguishable patches on various lines and surfaces in b.c. space. Most of these measurements were made at fairly high intensities and were confined to a particular class of just-distinguishable pairs, namely, those in which the members of each pair appeared equally bright (colour-limen type *). The observations on which the modified line-element of this paper is based were of a rather different nature, and are best introduced by considering the application of the Helmholtz element to another special class of just-distinguishable pairs of patches (liminal-increment type).

§2. APPLICATION OF THE HELMHOLTZ LINE-ELEMENT TO PAIRS OF JUST-DISTINGUISHABLE PATCHES OF LIMINAL INCREMENT TYPE

A pair is of this type if the physical stimulus in one patch is obtained from that of the other by the *addition* of a further stimulus. Consider the special

---

* Measurements of the hue limen amd the minimum perceptible colorimetric purity (Priest and Brickwedde, 1926) are of this type, as well as the more general colour limens recently determined by Wright (1941).

case when one patch is monochromatic, of wave-length μ and energy intensity W_μ, while the other is a mixture of wave-lengths μ and λ (λ in general different from μ), of energy intensities W_μ and U_λ respectively. The threshold condition (1) takes the form,

$$\left[\frac{U_\lambda r_\lambda'}{a+W_\mu r_\mu'}\right]^2 + \left[\frac{U_\lambda g'}{b+W_\mu g_\mu'}\right]^2 + \left[\frac{U_\lambda b_\lambda'}{c+W b_\mu'}\right]^2 = 3F^2, \quad \ldots\ldots(4)$$

where r_λ', g_λ', b_λ' are the fundamental response functions of the x, y, z or "red", "green", "blue" mechanisms respectively. Equation (4) may be written

$$\left(\frac{1}{U_\lambda}\right)^2 = [r_\lambda \eta(R_\mu W_\mu)]^2 + [g_\lambda \eta(G_\mu W_\mu)]^2 + [b_\lambda \eta(B_\mu W_\mu)]^2 \quad \ldots\vdots\ldots(5)$$

or

$$\left(\frac{1}{U_\lambda}\right)^2 = \left(\frac{1}{U_{\lambda r}}\right)^2 + \left(\frac{1}{U_{\lambda g}}\right)^2 + \left(\frac{1}{U_{\lambda b}}\right)^2, \quad \ldots\ldots(6)$$

where

$$\eta(p) = \left(\frac{1}{1+9p}\right)$$

and

$$r_\lambda = r_\lambda'/aF\sqrt{3}, \qquad R_\lambda = r_\lambda'/9a, \qquad \frac{1}{U_{\lambda r}} = r_\lambda \eta(R_\mu W_\mu),$$

$$g_\lambda = g_\lambda'/bF\sqrt{3}, \qquad G_\lambda = g_\lambda'/9b, \qquad \frac{1}{U_{\lambda g}} = g_\lambda \eta(G_\mu W_\mu), \quad \Bigg\} \quad \ldots\ldots(7)$$

$$b_\lambda = b_\lambda'/cF\sqrt{3}, \qquad B_\lambda = b_\lambda'/9c, \qquad \frac{1}{U_{\lambda b}} = b_\lambda \eta(B_\mu W_\mu).$$

$U_{\lambda r}$, $U_{\lambda g}$ amd $U_{\lambda b}$ may be regarded as the values which U_λ would assume if, in turn, each of the three mechanisms were acting alone. Since $\eta(0) = 1$, then, since $\frac{1}{U_{\lambda r}} = r_\lambda$ when $W_\mu = 0$, and r_λ is the reciprocal of the absolute threshold (in energy units) of the red mechanism for light of wave-length λ. Also, since $\eta(1) = 0\cdot1$, R_μ is the reciprocal of the energy intensity W_μ of light of wave-length μ required to raise $U_{\mu r}$, the difference threshold of the red mechanism, to ten times the absolute threshold ($1/r_\lambda$). Similar interpretations apply to g_λ, G_μ, b_λ, B_μ. For $\lambda = \mu$, we have

$$\frac{R_\lambda}{r_\lambda} = \frac{G_\lambda}{g_\lambda} = \frac{B_\lambda}{b^\lambda} = \frac{F\sqrt{3}}{9}. \quad \ldots\ldots(8)$$

For fixed λ and μ it is convenient to illustrate relation (5) by a plot of $\log(1/U_\lambda)$ against $\log W_\mu$, and to show plots of $\log(1/U_{\lambda r})$, $\log(1/U_{\lambda g})$ and $\log(1/U_{\lambda b})$ in the same diagram (figure 1). The latter curves (component curves) all have the same shape, but their respective positions in the diagram are determined by the values of r_λ, g_λ, b_λ (position on the ordinate axis) and R_μ, G_μ, B_μ (position on the abscissa axis). If $\mu = \lambda$, the component curves all tend asymptotically at high values of W_μ to the line : *ordinate* $= \log(1/F\sqrt{3}) - abscissa$, while the resultant curve ($\log 1/U_\lambda$) tends to the line: *ordinate* $= \log(1/F) - abscissa$. If $\mu \neq \lambda$, the component curves may intersect so that the resultant curve exhibits a more or less pronounced "change of law" as illustrated.

§3. BASIC MEASUREMENTS

Given the seven functions $\eta(p)$, $r_\lambda \ldots . B_\mu$ the resultant curve for any combination (λ, μ) can be drawn at once. An attempt at the converse problem of determining the seven functions from a number of experimental curves of log $(1/U_\lambda)$ against log W_μ has recently been made. The main points of this work (Stiles, 1939) will be briefly indicated. The present application was not in mind and the conditions of observation differed considerably from those implied in the foregoing discussion.

(*a*) The subject viewed, not two small juxtaposed patches but a large (10° diam.) patch of wave-length μ and energy intensity W_μ, at whose centre a small (1° square)
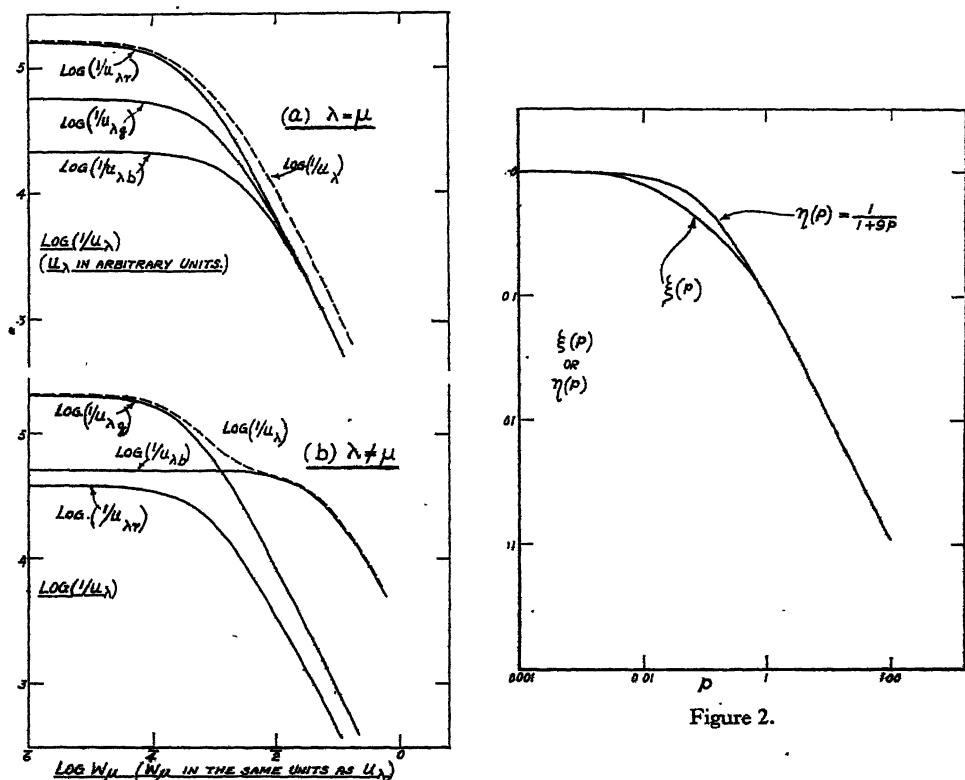
Figure 1.

Figure 2.

patch of wave-length λ and energy intensity U_λ' was applied in flashes (0·063 sec.) as an additional stimulus. Thus within the 1° square during the flash the energy intensity equalled $W_\mu + U_\lambda'$, while that of the surrounding field remained at W_μ throughout. The subject fixated the centre of the 10° patch, and great care was taken to ensure that the observations corresponded to foveal and cone (rod-free) vision of the additional stimulus. By repeated exposures of the latter at different energies U_λ', W_μ being kept constant, the critical intensity U_λ was determined at which the subject had a 50% chance of seeing the 1° patch.

(*b*) By measuring U_λ in this way for a series of values of W_μ from zero upwards, a curve of $\log(1/U_\lambda)$ against $\log W_\mu$ was obtained for various pairs of

wave-lengths λ and μ. These experimental curves corresponded to the broken-line curves in figure 1. For example, with $\lambda = 480$ mμ, $\mu = 540$ mμ, the experimental curve closely resembled the broken-line curve of figure 1 (b). It was the characteristic form of such experimental curves which first suggested that they were really resultant curves derivable from component curves of fixed shape. This view was supported by the fact that when λ was kept constant at 480 mμ and μ was varied, the experimental curve was modified as though (referring to figure 1 (b)) the component curves $\log(1/U_{\lambda g})$ and $\log(1/U_{\lambda b})$ were displaced parallel to the axis of $\log W_\mu$, while remaining in the same position along the axis of $\log(1/U_\lambda)$. On the other hand, when μ was kept constant and λ varied, the experimental curve was modified as though the component curves were displaced parallel to the axis of $\log(1/U_\lambda)$. These regularities in the results may be called the "displacement rules".

From nearly 100 curves of $\log(1/U_\lambda)$ against $\log W_\mu$ for the writer's left eye, it was concluded that all such curves for this eye could be regarded as resultants of three component curves of a common fixed shape whose positions in the diagram depended on λ and μ. The most satisfactory fixed shape to suit all the results was chosen by trial, and the positions in the diagram at which component curves of this shape had to be placed to yield a resultant curve agreeing with the experimental curve were determined for the various combinations of λ and μ studied. Actually no more than two component curves contributed to the experimental curve for any particular combination, but to cover all combinations in accordance with the displacement rules three components were necessary:

(c) The common shape of the component curves was represented as a function $\xi(p)$ (figure 2) defined so that $\xi(p) = 1$ for $p = 0$, $\xi(p) = 0 \cdot 1$ for $p = 1$. Thus $\xi(p)$ played the rôle of the function $\eta(p)$ above. It is apparent from figure 2 that the difference between the two functions is small and confined to values of p below 1. The expression for a particular component, say the "red" component, then took the form,

$$\log(1/U_{\lambda r}) = \log[r_\lambda \xi(R_\mu W_\mu)],$$

where for given λ and μ the constants r_λ and R_μ fix the position of the curve in the diagram. Conversely, when the position had been determined as explained above, the values of r_λ and R_μ were obtained. Although this method yielded values for each of the functions r_λ, R_μ, g_λ, etc., in a part only of the visible spectrum, the ranges were extended by assuming the ratios $\dfrac{R_\lambda}{r_\lambda}$, $\dfrac{G_\lambda}{g_\lambda}$, $\dfrac{B_\lambda}{b_\lambda}$ to be approximately independent of λ. This appeared to be nearly true in regions where both numerator and denominator functions could be followed. The tentative determinations of the six functions r_λ to B_μ are plotted as the continuous-line curves of figures 3 to 5. The absolute values depend on the unit in which energy intensity s expressed. Here and elsewhere in this paper, by *energy intensity* is to be understood the flux of radiation entering the eye, expressed in ergs per sec. per square degree of the light patch. The values given refer to the case when the pupil of entry at the eye is small and concentric with the natural pupil. Rather different values apply when the pupil is large or eccentric, on account of the directional properties of retinal sensitivity.

(d) On the assumption that $\dfrac{R_\lambda}{r_\lambda}$, $\dfrac{G_\lambda}{g_\lambda}$ and $\dfrac{B_\lambda}{b_\lambda}$ are constants independent of λ,

the values of these ratios were found to be (very approximately) in the proportion

$$\frac{R_\lambda}{r_\lambda} : \frac{G_\lambda}{g_\lambda} : \frac{B_\lambda}{b_\lambda} = 0\cdot 78 : 1 : 4\cdot 46.$$

(e) The resultant curve was related to the component curves by the following rough rule : for a given $\log W_\mu$, $\log 1/U_\lambda \dots$ equals the greatest of $\log (1/U_{\lambda r})$ $\log (1/U_{\lambda g})$ and $\log (1/U_{\lambda b})$, except when the greatest and next greatest have nearly the same value. In the latter case $\log (1/U_\lambda)$ exceeds the greatest by a small amount, of the order of $0\cdot 1$.
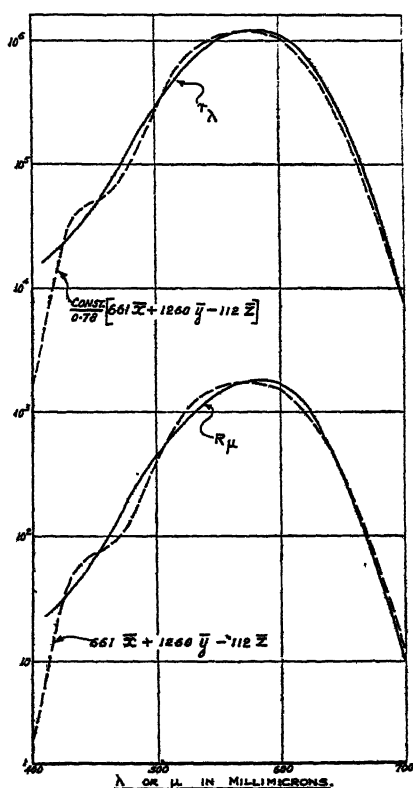
Figure 3.
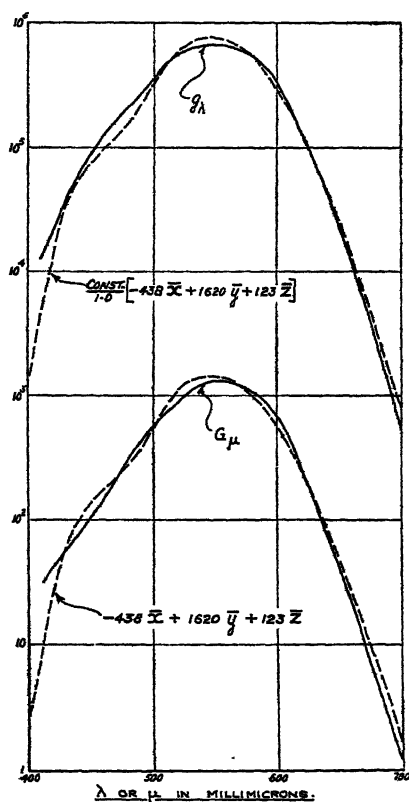
Figure 4.

(f) A breakdown in the analysis of the results on the above lines occurred when the 10° patch was orange or red and of very high intensity (about 30,000 photons) and when the additional stimulus was blue (*limited conditioning effect of red light on the blue mechanism*). Colour-match equations also break down after adaptation to intensities of this order (Wright, 1936).

Setting aside the effect noted in (f), the experimental results differ from the predictions of the line-element and associated primaries of Helmholtz in the following respects :—

(i) The function $\eta(p)$ is replaced by a slightly different function $\xi(q)$.

(ii) The spectral sensitivity functions have the same general form as König's *Grundempfindungen*.

(iii) The ratios $\dfrac{R_\lambda}{r_\lambda}$, $\dfrac{G_\lambda}{g_\lambda}$, $\dfrac{B_\lambda}{b_\lambda}$ are not equal: $\dfrac{B_\lambda}{b_\lambda}$ is materially greater than the other two. It follows that for $\lambda = \mu$, the three component curves in diagrams such as figure 1 no longer tend to a common line at high intensities.

(iv) The effects of the three mechanisms add up, if anything, to a smaller extent than required by the sum of squares relation (6).
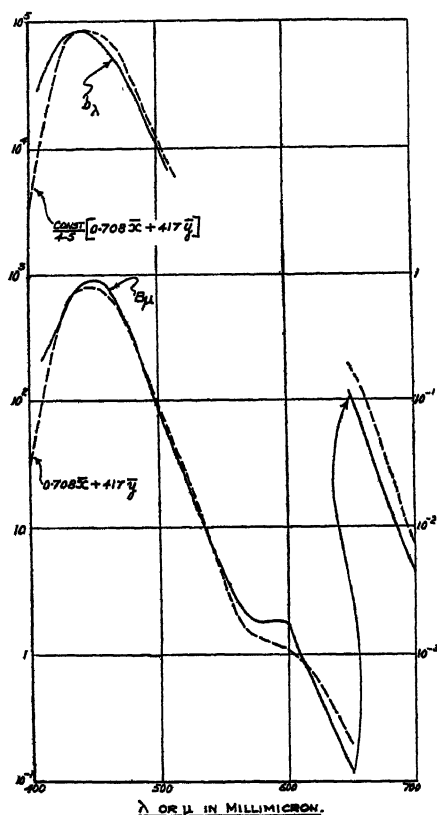
Figure 5.

For $\lambda = \mu$, the ratios $\dfrac{U_{\lambda r}}{W_\lambda}$, $\dfrac{U_{\lambda g}}{W_\lambda}$, $\dfrac{U_{\lambda b}}{W_\lambda}$ are the Fechner fractions * of the respective mechanisms. Since $\xi(p)$ tends to $\dfrac{1}{9p}$ as p becomes large, it is clear that $\dfrac{U_{\lambda r}}{W} = \dfrac{1}{W_\lambda r_r \xi(R_\lambda W_\lambda)}$ approaches $9\dfrac{R_\lambda}{r_\lambda}$ at high intensities W_λ. Thus the ratios $\dfrac{R_\lambda}{r_\lambda}$, $\dfrac{G_\lambda}{g_\lambda}$, $\dfrac{B_\lambda}{b_\lambda}$ are proportional respectively to the *limiting Fechner* fractions of the

* Historically it would be more accurate to call these ratios " Weber fractions ", but it has seemed preferable to follow the more usual practice.

three mechanisms, and (iii) implies that, contrary to the assumption of Helmholtz, the different mechanisms have different limiting Fechner fractions.

The precision of threshold measurements is low, and during a long investigation the properties of the eye may undergo systematic changes. There can be no question of a precise determination of the unknown functions by the methods used. All that can be claimed is that the derived functions reproduce the main features of the experimental curves.

§ 4. MODIFIED LINE-ELEMENT

Following the hints given by the basic measurements, we may construct a tentative line-element for application to any pair of just-distinguishable light patches, not necessarily of liminal increment type. Suppose the patches P' and P have absolute energy distributions $W_\lambda' d\lambda$ and $\cdot W_\lambda d\lambda$ respectively. Put

$$\zeta(p) = 9\xi(p), \qquad R = \int W_\lambda R_\lambda d\lambda, \qquad \delta R = \int (W_\lambda' - W_\lambda) R_\lambda d\lambda \quad \ldots\ldots(10)$$

and define G, δG, B, δB similarly, where $\xi(p)$, R_λ, G_λ, B_λ are the *experimental functions** plotted in figures 2 to 5. Introduce the quantities ρ, γ, β proportional to the limiting Fechner fractions of the three mechanisms,

$$\rho : \gamma : \beta = \frac{R_\lambda}{r_\lambda} : \frac{G_\lambda}{g_\lambda} : \frac{B_\lambda}{b_\lambda} = 0 \cdot 78 : 1 : 4 \cdot 46, \qquad \ldots\ldots(11)$$

and satisfying

$$\frac{1}{\rho^2} + \frac{1}{\gamma^2} + \frac{1}{\beta^2} = 1, \qquad \ldots\ldots(12)$$

so that $\dfrac{1}{\rho^2} = 0 \cdot 612$, $\dfrac{1}{\gamma^2} = 0 \cdot 369$, $\dfrac{1}{\beta^2} = 0 \cdot 0185$. The proposed line-element is then

$$\delta s^2 \equiv \left[\frac{\delta R}{\rho} \zeta(R)\right]^2 + \left[\frac{\delta G}{\gamma} \zeta(G)\right]^2 + \left[\frac{\delta B}{\beta} \zeta(B)\right]^2 = F^2, \quad \ldots\ldots(13)$$

which reduces at high intensities to

$$\delta s^2 \equiv \left[\frac{1}{\rho} \frac{\delta R}{R}\right]^2 + \left[\frac{1}{\gamma} \frac{\delta G}{G}\right]^2 + \left[\frac{1}{\beta} \frac{\delta B}{B}\right]^2 = F^2, \qquad \ldots\ldots(14)$$

where F is a constant.

If the two patches have the same *relative* energy distributions, their intensities may be specified by their total energies T' and T (or by their total energies weighted according to any function of wave-length, such as the visibility curve). Their Fechner fraction is then given by

$$\frac{\delta T}{T} \equiv \frac{T' - T}{T} = \frac{\delta R}{R} = \frac{\delta G}{G} = \frac{\delta B}{B}.$$

Thus, by (14) and (12), $\dfrac{\delta T}{T} = F$ at high intensities, or F is the limiting Fechner fraction for any pair of patches of the same relative energy distribution whatever form that may take. The factor 3 introduced into (1) gives F a similar meaning for the Helmholtz line-element. Experimentally, the value of F depends on the

* The definitions of R_λ, G_λ, B_λ are modified later.

precise conditions of observation. It is round about 0·01 for the conditions normally used in determining colour limens, and this value will be used in the calculations which follow.

For just distinguishable patches of liminal increment type, δR, δG and δB are all positive or all negative. In applying the line-element generally, it is assumed that the signs of δR, δG, δB are immaterial, so that, for example, slight excess stimulations of the red and green mechanisms make the same contribution to the limen whether the excesses occur both in the same or one in each patch.

§5. COLOUR MATCHING

Light patches whose representative points P′ in b.c. space lie within the small ellipsoid centred on P and defined by (13) are indistinguishable from, and therefore in colour match with, P. In colour-matching measurements the objective is to bring the variable colour to the centre of this ellipsoid, and the condition for an ideal match may be taken to be $\delta R = \delta G = \delta B = 0$, or $\int W_\lambda' R_\lambda \, d\lambda = \int W_\lambda R_\lambda \, d\lambda$, and two similar equations.* R_λ, G_λ, B_λ were, in fact, obtained by a method completely independent of colour-matching, but their introduction into the line-element relates them at once to the distribution coefficients $\bar{x}$, $\bar{y}$, $\bar{z}$ derivable from colour-matching measurements. In terms of these coefficients, the conditions for a colour match are

$$\int W_\lambda' \bar{x} \, d\lambda = \int W_\lambda \bar{x} \, d\lambda$$

and two similar equations. Thus R_λ, G_λ, B_λ should be linear forms in $\bar{x}$, $\bar{y}$, $\bar{z}$. The distribution coefficients for the writer's eye (a normal trichromat) have not been determined, but using the coefficients of the C.I.E. Standard Observer in the Standard Reference System (C.I.E. 1931) an approximate linear representation of R_λ, G_λ, B_λ is obtained as follows :—

$$
\begin{aligned}
(R_\lambda) & \quad 6\cdot61 \cdot 10^2 \, \bar{x} + 1\cdot26 \cdot 10^3 \bar{y} - 1\cdot12 \cdot 10^2 \bar{z}, \\
(G_\lambda) & \quad -4\cdot38 \cdot 10^2 \, \bar{x} + 1\cdot62 \cdot 10^3 \bar{y} + 1\cdot23 \cdot 10^2 \bar{z}, \\
(B_\lambda) & \quad 7\cdot08 \cdot 10^{-1} \bar{x} + 0 \cdot \bar{y} \quad\quad + 4\cdot17 \cdot 10^2 \bar{z},
\end{aligned}
\quad \dots\dots (15)
$$

These linear forms are plotted as the broken lines in figures 3 to 5. The agreement between the broken and continuous curves is fairly satisfactory when it is recalled (*a*) that the experimental ·uncertainties in determining R_λ, G_λ, B_λ are considerable—much greater than for $\bar{x}$, $\bar{y}$, $\bar{z}$; (*b*) that results for a single eye and for an average eye are being compared. The basic measurements (§ 3) could be represented almost as well by using the expressions (15) in place of R_λ, G_λ, B_λ, and by making the same substitutions in the definitions of R, δR, etc., the line element would lead automatically tó the correct colour-matching equations for the C.I.E. standard observer. From now on, therefore, R_λ, G_λ, B_λ are supposed defined by the expressions (15). We are, of course, still dependent on the basic measurements for the coefficients in (15) and for the relative values of ρ, γ, β.

* Even if an ideal colour match is not achieved, (13) ensures that these equations shall be approximately true, i.e. δR will be small compared with R, and so on.

§6. STEP-BY-STEP VISIBILITY CURVE

In the experimental determination of the visibility curve by the step-by-step method, the two juxtaposed patches have slightly different colours and the intensity of one is varied until the "difference" between the two patches is a minimum. If the colour difference is appropriately chosen, the two patches are just indistinguishable at this minimum setting of the intensity. The visibility curve is derived from the line-element (14) by a corresponding procedure. The unit co-ordinates (in the fundamental system) of the two patches, i', j', k' and i, j, k, are required, where

$$i' = \frac{R'}{R'+G'+B'} = \frac{R'}{\Sigma'}, \qquad i = \frac{R}{R+G+B} = \frac{R}{\Sigma}, \qquad \ldots\ldots(16)$$

and j', k' ; j, k are similarly defined.

In terms of these, the line-element (13) takes the form

$$\delta s^2 \equiv \left[\frac{\Sigma' i' - \Sigma i}{\rho} \zeta(i) \right]^2 + \text{two similar terms.} \qquad \ldots\ldots(17)$$

By varying Σ', keeping i', j', k' constant, the intensity of patch P' is varied without altering its colour. The minimum value of δs^2 under such a variation is reached when

$$0 = \frac{1}{\rho^2} [2i'(\Sigma' i' - \Sigma i)\zeta(\Sigma i)]^2 + \text{two similar terms,}$$

or, reverting to the earlier notation and retaining only first-order quantities, when

$$0 = \frac{\delta R}{R} \left[\frac{R}{\rho} \zeta(R) \right]^2 + \frac{\delta G}{G} \left[\frac{G}{\gamma} \zeta(G) \right]^2 + \frac{\delta B}{B} \left[\frac{B}{\beta} \zeta(B) \right]^2. \qquad \ldots\ldots(18)$$

Strictly, the colour difference of the two patches should be chosen so that at the minimum setting $\delta s = F$, but it suffices both in the theory and in the experimental method for this condition merely to be approached.

The equality-of-brightness condition (18) applies to any pair of patches of nearly the same colour. In the special case of monochromatic patches of wave-lengths λ, $\lambda + \delta\lambda$ and energy intensities W_λ, $W_\lambda + \delta W_\lambda$, (18) becomes

$$-\frac{\delta W_\lambda}{W_\lambda} = \delta\lambda \left[\frac{C_r^2}{R_\lambda} \frac{dR_\lambda}{d\lambda} + \frac{C_g^2}{G_\lambda} \frac{dG_\lambda}{d\lambda} + \frac{C_b^2}{B_\lambda} \frac{dB_\lambda}{d\lambda} \right], \qquad \ldots\ldots(19)$$

where

$$C_r^2 = \frac{\left[\frac{R}{\rho} \zeta(R) \right]^2}{\left[\frac{R}{\rho} \zeta(R) \right]^2 + \left[\frac{G}{\gamma} \zeta(G) \right]^2 + \left[\frac{B}{\beta} \zeta(B) \right]^2} \qquad \ldots\ldots(20)$$

and C_g^2, C_b^2 are similarly defined. By successive applications of (19) we can determine the energy intensities of a series of monochromatic patches whose wave-lengths increase in small steps from the blue to the red, and each of which matches in brightness its next neighbours in the series. The reciprocal of the patch energy plotted against wave-length represents a step-by-step visibility curve whose shape will depend in general on the intensity level, specified, say, by the

value of W_λ at $\lambda = 555$ mμ. At sufficiently high intensity levels $C_r{}^2$, $C_g{}^2$, $C_b{}^2$ tend to $1/\rho^2$, $1/\gamma^2$, $1/\beta^2$ respectively, and the visibility curve approaches the limiting form

$$V_\lambda = \frac{\text{const.}}{W_\lambda} = \text{const. } R_\lambda{}^{1/\rho^2}\, G_\lambda{}^{1/\gamma^2}\, B_\lambda{}^{1/\beta^2}. \qquad \ldots \ldots (21)$$

As usual, the constant is adjusted so that V_λ has the maximum value unity.

The circle points in figure 6 represent relative visibilities calculated from the expression (21). Except in the blue, these points lie very close to the full-line curve, which is a plot of the C.I.E. visibility function. The discrepancy in the blue is greatly reduced if the comparison is made with the mean visibility curve obtained by Gibson and Tyndall (1923) for 52 observers (38 on ends) using the step-by-step method.* In the several investigations of the visibility curve in which large groups of observers have been used, the variations between
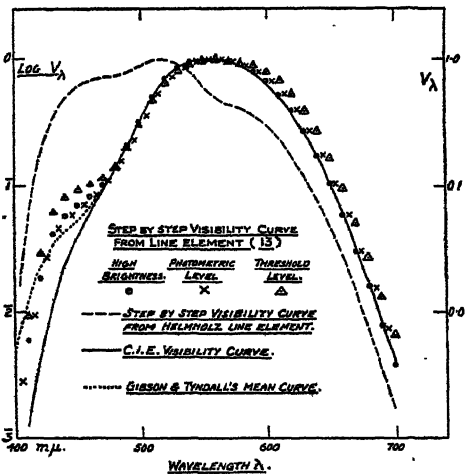
Figure 6. Step-by-step visibility curves.
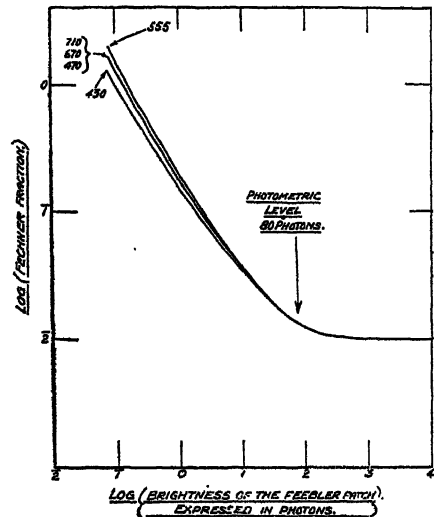
Figure 7. Fechner fraction curves.

observers are markedly greater in the blue (say 455 mμ) than in the red (say 655 mμ). On this general ground any theoretical derivation of the visibility curve may be expected to meet greater difficulties in the blue than in the red.

The broken curve in figure 6 is the visibility function for the Helmholtz line-element, but assuming the response functions (15). It is derived from (21) by putting

$$\frac{1}{\rho^2} = \frac{1}{\gamma^2} = \frac{1}{\beta^2} = \frac{1}{3}.$$

Equally unacceptable curves are obtained by applying to the Helmholtz element König's *Grundemfindungen* or the double-peak response curves of Helmholtz. The Hecht response curves (Hecht, 1932), each of which is a fair approximation to the visibility curve, would make the Helmholtz or the modified line-element consistent with the visibility curve, but would render it incapable of explaining the basic measurements (§ 4).

* I am indebted to Prof. Hecht for drawing my attention to this point. He also points out that the mean curve of Coblenz and Emerson (1917) shows a hump in the blue, and that a protanope studied by Hecht and Shlaer gave a curve similar in this respect to a normal observer.

The standard photometric brightness level at which the C.I.E. visibility curve applies may be taken as 25 equivalent metre candles seen through an eye pupil of 10 sq. mm. area, or 79·5 photons. A monochromatic patch of wavelength 555 mμ has this brightness when its energy intensity W_λ equals $3·66 \times 10^{-4}$ erg per sec. per square degree. The limiting form (21) of the derived visibility curve is barely reached at this value of W_{555}, and the curve must be derived from (19) by a method of successive approximation. The results of the calculation are shown as the cross-points in figure 6. The agreement with the C.I.E. curve is substantial, although in the red it is rather less close than for the high-intensity curve. As the intensity is reduced below the photometric level, the derived curve undergoes some further modification, to assume finally the form shown by the triangle-points of figure 6, when the absolute threshold is reached. The second limiting form (22) can be obtained from (19) or directly from the line-element (13) by using the principle that at the absolute threshold all light patches are equally bright :

$$V_\lambda = \text{const.} \sqrt{\left(\frac{R_\lambda}{\rho}\right)^2 + \left(\frac{G_\lambda}{\gamma}\right)^2 + \left(\frac{B_\lambda}{\beta}\right)^2}. \qquad \ldots\ldots (22)$$

The main effects of lowering the intensity level are to increase slightly the visibilities in the orange and red and in the blue. It must be emphasized that the line-element (13) is intended to represent foveal and strictly cone vision. The small changes in the visibility curve just considered are quite distinct from the very much larger changes (Purkinje effect) which occur when, by the use of a large or eccentrically fixated matching field, the rods participate in the matching process.

No determinations of the foveal cone visibility curve by the step-by-step method have been made at low intensities. Walters and Wright (1943) used a fixed red comparison patch and worked with the full-colour difference. They found that at low intensities the visibility curve showed a slight shift of the maximum towards the blue, which they think may have resulted from some admixture of rod vision, and a " lump " in the curve at about 600 mμ, which they attribute to a tendency of the red mechanism to retain its sensitivity over a greater intensity range than the green or blue mechanisms.

Relation (18) is the differential of the function

$$P(R, G, B) = P_r(R) + P_g(G) + P_b(B),$$

where

$$P_r(R) = \int^R \frac{dR}{R}\left[\frac{R\zeta(R)}{\rho}\right]^2,$$

and P_g, P_b are similarly defined. All lights (not necessarily monochromatic) for which $P(R, G, B)$ has the same value can be linked in small steps by series of intermediate lights, adjacent lights in any series having the same brightness. We may say that all such lights have the same " small-step " brightness. But P_r, P_g and P_b are not linear functions of their arguments, so that " small-step " brightness is not additive. To estimate the magnitude of the breakdown in the additive law, we will use for $P(R, G, B)$ the form to which it reduces at high intensities, namely,

$$\log_e [R^{1/\rho^2} G^{1/\gamma^2} B^{1/\beta^2}].$$

If lights 1 and 2 have the same small-step brightness,

$$R_1{}^{1/e^2} G_1{}^{1/\gamma^2} B_1{}^{1/\beta^2} = R_2{}^{1/e^2} G_2{}^{1/\gamma^2} B_2{}^{1/\beta^2},$$

and the equality of brightness still holds if their intensities are changed by the same factor, α say, since

$$(R_1\alpha)^{1/e^2} (G_1\alpha)^{1/\gamma^2} (B_1\alpha)^{1/\beta^2} = (R_2\alpha)^{1/e^2} (G_2\alpha)^{1/\gamma^2} (B_2\alpha)^{1/\beta^2}.$$

But if lights 1 and 2 each match light 3 in brightness, the mixture of 1 and 2 will not in general match light 3 increased to double its original intensity, since

$$(R_1+R_2)^{1/e^2}(G_1+G_2)^{1/\gamma^2}(B_1+B_2)^{1/\beta^2} = (2R_3)^{1/e^2}(2G_3)^{1/\gamma^2}(2B_3)^{1/\beta^2}$$

is not in general true. The discrepancy may be expected to be greatest when the colours of 1 and 2 are most widely different, and it is estimated that about the worst case arises with monochromatic lights in the blue (470 mμ) and red (680 mμ). The colour of light 3 is immaterial. Calculation for this case shows that

$$(R_1+R_2)^{1/e^2}(G_1+G_2)^{1/\gamma^2}(B_1+B_2)^{1/\beta^2} = (2 \cdot 63R_3)^{1/e^2}(2 \cdot 63G_3)^{1/\gamma^2}(2 \cdot 63G_3)^{1/\beta^2}.$$

The breakdown in additivity corresponds to the difference between 2 and 2·63, or about 27%. As a second example, take a white of energy distribution $W_\lambda d\lambda$ appropriate to a black body of temperature 2060°K. and compare its small-step brightness with the sum of the small-step brightnesses of its component wavelengths. Additivity would require

$$\int R_\lambda{}^{1/e^2} G_\lambda{}^{1/\gamma^2} B_\lambda{}^{1/\beta^2} W_\lambda d\lambda = \left\{\int W_\lambda R_\lambda d\lambda\right\}^{1/e^2} \left\{\int W_\lambda G_\lambda d\lambda\right\}^{1/\gamma^2} \left\{\int W_\lambda B_\lambda d\lambda\right\}^{1/\beta^2},$$

but calculation shows that the equation is in error by about 8%. This second case is the one used by Ives to test the additivity of brightness as assessed in the flicker photometer. He found additivity held good within the experimental error (1 or 2%). No similar check of additivity for small-step brightness has been made, and it cannot be said with certainty that discrepancies of the kind indicated by the line-element would not be observed, although there is no positive evidence for them. The results of brightness matching by flicker photometry, like those obtained by direct-comparison photometry with full colour difference, cannot be derived from the line-element without additional assumptions and will not be discussed.

§7. FECHNER FRACTION CURVES

Although the modified line-element assumes that the red, green and blue mechanisms have different limiting Fechner fractions at high intensities, it still leads to a common value F for the limiting Fechner fraction of a pair of just-distinguishable and similarly-coloured light patches, whatever that colour may be. This is in accordance with König and Brodhun's classic measurements of the Fechner fractions of white and monochromatic light. These measurements were made with large patches, each $3° \times 4\frac{1}{2}°$, and it is only at the higher intensities studied that the results can be accepted as referring to cone vision. The wide divergence in the observed values of the Fechner fraction at lower intensities is caused by the intrusion of the rods. The change with intensity in the Fechner fraction of monochromatic patches as predicted by the modified line-element (13)

is shown in figure 7. The Fechner fraction, plotted as ordinate, is the just-distinguishable difference in patch intensity divided by the intensity of the *feebler* patch, and the intensity level, plotted as abscissa, is the intensity of the feebler patch. For $\lambda = 555\,m\mu$, this intensity is expressed in photons and for other wave-lengths in units such that all colours have the same brightness when compared on the basis of the calculated step-by-step visibility curve for the photometric level (figure 6). The Fechner fraction curves for different wave-lengths are not very different, but for the extreme blue the Fechner fraction has a rather smaller value at low intensities than for other colours.

§8. HUE LIMEN

The hue limen, or the difference of wave-length of two just-distinguishable monochromatic patches of equal brightness, is obtained from relations (13) and (18) by the substitutions.

$$R = W_\lambda R_\lambda, \quad \delta R = R_\lambda \delta W_\lambda + W_\lambda \frac{dR_\lambda}{d\lambda} \delta\lambda, \text{ etc.}$$

On eliminating W_λ, the equation for the hue limen $\delta\lambda$ takes the form :

$$F^2 = \delta\lambda^2 \left[\left\{ \frac{C_r^2 - 1}{R_\lambda} \frac{dR_\lambda}{d\lambda} + \frac{C_g^2}{G_\lambda} \frac{dG_\lambda}{d\lambda} + \frac{C_b^2}{B_\lambda} \frac{dB_\lambda}{d\lambda} \right\}^2 \left\{ \frac{R}{\rho} \zeta(R) \right\}^2 + \text{two similar terms} \right]$$

$$\ldots\ldots(23)$$

At high intensities this reduces to

$$F^2 = \delta\lambda^2 \left[\left\{ \frac{1/\rho^2 - 1}{R_\lambda} \frac{dR_\lambda}{a\lambda} + \frac{1/\gamma^2}{G_\lambda} \frac{dG_\lambda}{d\lambda} + \frac{1/\beta^2}{B_\lambda} \frac{dB}{d\lambda} \right\}^2 1/\rho^2 + \text{two similar terms} \right].$$

$$\ldots\ldots(24)$$

The equation used by Helmholtz is the special case of (24), when ρ, γ, β are made equal.

The values of the hue limen through the spectrum have been calculated from (23), assuming a common brightness for all colours on the basis of the calculated step-by-step visibility curve for the photometric level (figure 6) and an actual brightness of 80 photons for $\lambda = 550\ m\mu$.

Table 1. Values of hue limen at the photometric level derived from the modified line element

λ (mμ)	$\delta\lambda$ (mμ)	λ (mμ)	$\delta\lambda$ (mμ)	λ (mμ)	$\delta\lambda$ (mμ)
410	5·7	510	2·06	610	1·43
420	3·9	520	2·74	620	1·88
430	2·22	530	2·88	630	2·42
440	1·75	540	2·87	640	3·2
450	1·41	550	2·62	650	4·7
460	1·63	560	2·03	660	6·8
470	1·14	570	1·72	670	11
480	0·82	580	1·44	680	18
490	0·97	590	1·27	690	29
500	1·45	600	1·26		

The calculated values are plotted in figure 8 together with five experimental curves obtained by Wright and Pitt (1934) at an average intensity level (480 to

650 mμ) of 70 photons. An outstanding difference between the theoretical and experimental curves is that in the former the limens in the blue are relatively low.
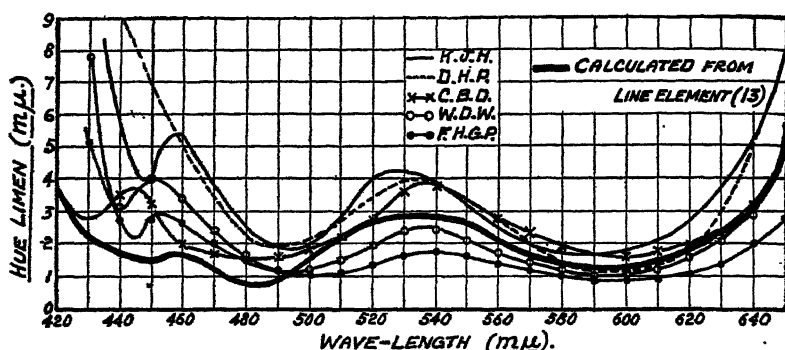
Figure 8. Observed (Wright and Pitt, 1934) and calculated values of the hue limen.

In the determination of the experimental curves it was necessary to use lower intensities in the extreme blue, the total range being about 20 to 1. This may account for some of the difference.

§9. GENERAL COLOUR LIMEN

If two patches P and P', not necessarily monochromatic, have the same brightness and are just distinguishable, they define a general colour limen. The general colour limens for a given common brightness level are conveniently displayed in a C.I.E. (x, y) chart (a) by marking pairs of just-distinguishable colours on various lines joining spectrum colours to other spectrum colours or to purples (Wright, 1941), or (b) by drawing small closed curves (ellipses) about various points in the chart, the colours corresponding to all points on each such curve being just distinguishable from the central colour. The modified line-element will now be applied to calculate the g.c.l. (general colour limen) ellipses for the 25 colours used in the recent investigation by MacAdam (1942). Using an ingenious optical system, MacAdam arranged that, by turning a single control knob, the colour of a light patch was varied along a line in the C.I.E. chart while its brightness (C.I.E.) was automatically held constant. Repeated colour matches were made with a second juxtaposed patch of the same brightness and of a fixed colour on the given line. The standard deviations of the settings were determined, a setting being specified by the distance in the chart between the variable and fixed points. The standard deviation is closely related to, and may be taken as proportional to, the diameter of the g.c.l. ellipse at the fixed colour in the direction of the given line. MacAdam determined for one subject the complete ellipses for 25 fixed colours distributed fairly evenly over the domain between the spectrum locus and the line of purples.

Denoting by $t_{mn}(m, n = 1, 2, 3)$ the coefficients in the forms (15), the R co-ordinate of a patch of absolute energy distribution $W_\lambda d\lambda$ is given by

$$R = t_{11}x' + t_{12}y' + t_{13}z', \qquad \ldots\ldots(25)$$

where $\qquad x' = \int W_\lambda \bar{x} \, d\lambda, \qquad y' = \int W_\lambda \bar{y} \, d\lambda, \qquad z' = \int W_\lambda \bar{z} d\lambda,$

and similar equations hold for G and B. The C.I.E. unit co-ordinates x, y, z are then $x=x'/w'$, $y=y'/w'$, $z=z'/w'$, where $w'=x'+y'+z'$. The quantity y' is proportional to the C.I.E. brightness of the patch and, in fact, recalling the units in which $W_\lambda d\lambda$ is expressed, we have :

$$\text{patch brightness in photons} = 2\cdot19 \cdot 10^5 \int W_\lambda \bar{y}\, d\lambda,$$

where the mechanical equivalent of light has been taken as $0\cdot0015$ watts per lumen. MacAdam worked with a C.I.E. brightness of 15 millilamberts seen through a pupil of $2\cdot6$ mm. diameter, that is, with a brightness of 271 photons. The calculation will be applied, therefore, to the case, -

$$\int W_\lambda \bar{y}\, d\lambda = y' = w'y = \frac{271}{2\cdot19\cdot10^5} = 1\cdot24 \cdot 10^{-3}.$$

The R co-ordinate of the just-distinguishable patch is $R+\delta R$, where

$$\delta R = t_{11}\delta x' + t_{12}\delta y' + t_{13}\delta z',$$

and similar equations hold for δG and δB. In MacAdam's observations $\delta y' = 0$. The differentials $\delta x'$, $\delta z'$, expressed in terms of dx, dy, are then :

$$\delta x' = \frac{y'}{y}\left\{\delta x - \frac{x}{y}\delta y\right\},$$

$$\delta z' = \frac{y'}{y}\left\{\frac{x-1}{y}\delta y - \delta x\right\}.$$

Thus,

$$\delta R = \delta x \cdot \frac{y'}{y}(t_{11}-t_{13}) + \delta y \cdot \frac{y'}{y^2}(x\overline{t_{13}-t_{11}}-t_{13}),$$

$$R = x \cdot \frac{y' t_{11}}{y} + y' t_{12} + z \cdot \frac{y'}{y} t_{13}, \qquad\qquad \dots\dots (26)$$

and, inserting these expressions, the line-element (13) takes the form

$$F^2 = \delta x^2\left[\left\{\frac{p_{11}}{\rho}\zeta(R)\right\}^2 + \left\{\frac{p_{21}}{\gamma}\zeta(G)\right\}^2 + \left\{\frac{p_{31}}{\beta}\zeta(B)\right\}^2\right]$$

$$+ 2\delta x\,\delta y\left[p_{11}p_{12}\left\{\frac{\zeta(R)}{\rho}\right\}^2 + p_{21}p_{22}\left\{\frac{\zeta(G)}{\gamma}\right\}^2 + p_{31}p_{32}\left\{\frac{\zeta(B)}{\beta}\right\}^2\right]$$

$$+ \delta y^2\left[\left\{\frac{p_{12}}{\rho}\zeta(R)\right\}^2 + \left\{\frac{p_{22}}{\gamma}\zeta(G)\right\}^2 + \left\{\frac{p_{32}}{\beta}\zeta(B)\right\}^2\right], \qquad \dots\dots (27)$$

where

$$p_{11} = \frac{y'}{y}(t_{11}-t_{13}), \qquad p_{12} = \frac{y'}{y^2}(x\overline{t_{13}-t_{11}}-t_{13}).$$

$$p_{21} = \frac{y'}{y}(t_{21}-t_{23}), \qquad p_{22} = \frac{y'}{y^2}(x\overline{t_{23}-t_{21}}-t_{23}),$$

$$p_{31} = \frac{y'}{y}(t_{31}-t_{32}), \qquad p_{32} = \frac{y'}{y^2}(x\overline{t_{33}-t_{31}}-t_{33}),$$

For a given colour x, y, the g.c.l. ellipse is completely determined by the equation

(27). The semi-axes and orientation of the ellipse have been calculated for MacAdam's 25 colours and are as follows:—

Table 2

Colour MacAdam's figure number	C.I.E. co-ordinates			Semi-axes		Angle between major axis and axis of x (degrees)
	x	y	z	Major	Minor	
23	0·160	0·057	0·783	$0·235 . 10^{-2}$	$0·139 . 10^{-2}$	85
24	0·187	0·118	0·695	0·62	0·285	75
25	0·253	0·125	0·622	0·71	0·272	53
26	0·150	0·680	0·170	2·84	1·26	106
27	0·131	0·521	0·348	2·60	1·04	101
28	0·212	0·550	0·238	2·37	1·13	97
29	0·258	0·450	0·292	1·90	0·97	86
30	0·152	0·365	0·483	1·86	0·76	96
31	0·280	0·385	0·335	1·67	0·85	78
32	0·380	0·498	0·122	1·53	1·04	83
33	0·160	0·200	0·640	1·10	0·44	92
34	0·228	0·250	0·522	1·24	0·57	76
35	0·305	0·323	0·372	1·43	0·72	68
36	0·385	0·393	0·222	1·47	0·83	63
37	0·472	0·399	0·129	1·23	0·78	54
38	0·517	0·350	0·123	1·14	0·61	44
39	0·475	0·300	0·225	1·27	0·52	43
40	0·510	0·236	0·254	1·17	0·31	35
41	0·596	0·283	0·121	1·02	0·35	33
42	0·344	0·284	0·372	1·30	0·60	57
43	0·390	0·237	0·373	1·18	0·44	45
44	0·441	0·198	0·361	1·13	0·284	36
45	0·278	0·223	0·499	1·14	0·49	62
46	0·300	0·163	0·537	0·91	0·33	50
47	0·365	0·153	0·482	0·99	0·24	39

In figures 9 and 10 the calculated and MacAdam's ellipses are plotted in the C.I.E. chart. There is some similarity. The orientations of the ellipses correspond rather well (figure 11), and their areas show a correlation (figure 12). The computed areas exceed MacAdam's on the average by a factor of 20·7 ($4·5^2$), but MacAdam's refer to standard deviations, which he estimates to be about one-third the corresponding limens. The axis ratios of the ellipses cover a range of about 2·5 to 1 and have average values 2·3 (calculated) and 2·9 (MacAdam's), but they show no correlation (figure 13).

MacAdam reports a limited number of observations for a second subject. The ratio of the g.c.l.s of the two subjects varies, for different points and directions in the chart, over a range of about 0·8 to 2·5. For colour No. 35 sufficient results are given to construct the ellipse for the second subject (figure 14*), and in table 3 the characteristics of the ellipses of the two subjects are compared.

* The plotted points in this figure were read off the smooth curves of MacAdam's figures 8 to 19, and there will be some reading error.
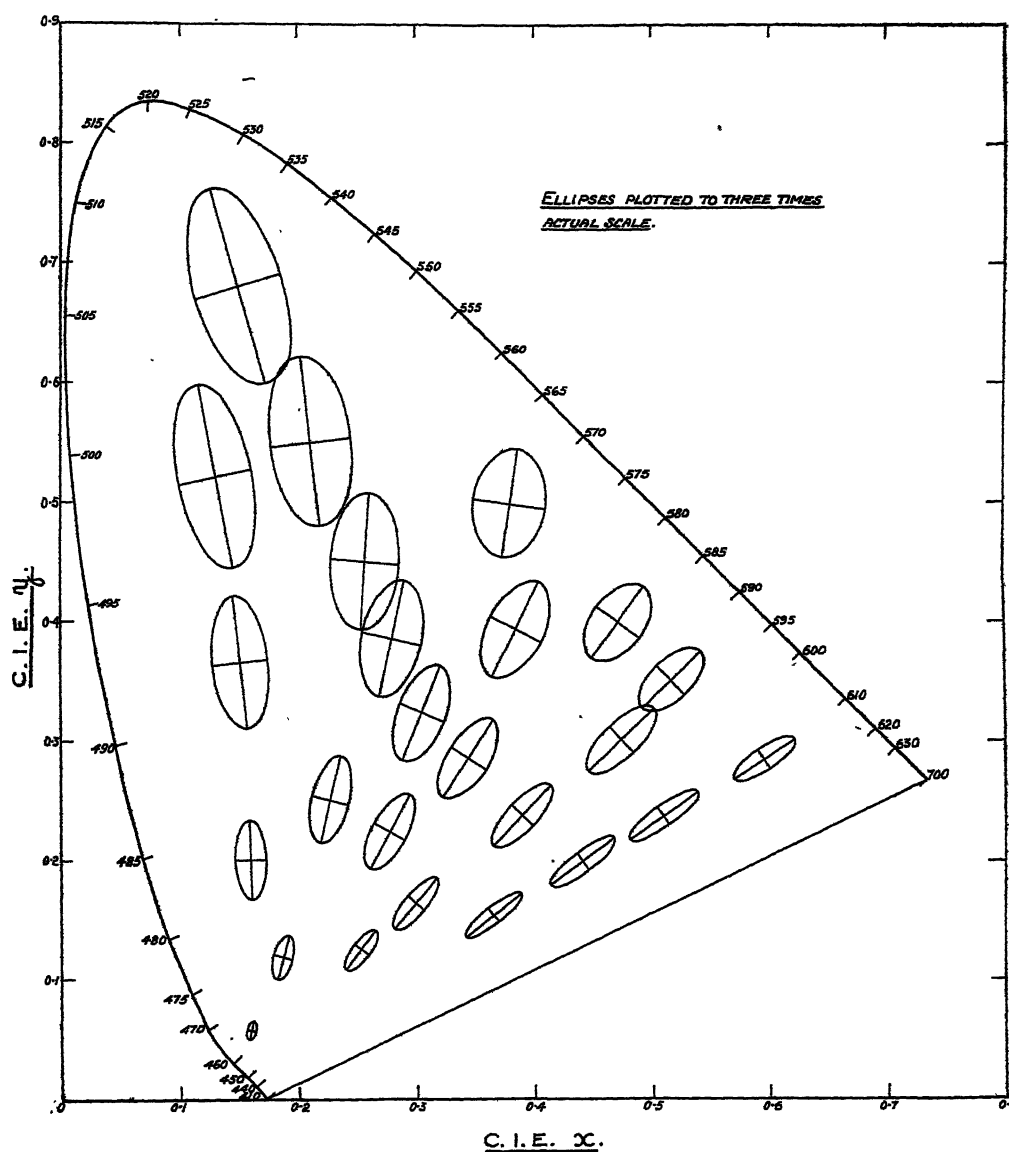
Figure 9. G.C.L. ellipses calculated from the line-element (13).

Table 3

	Major semi-axis	Minor semi-axis	Ratio of axes	Area of ellipse	Inclination of major axis to axis of x (degrees)
Principal subject (P.G.N.)	$0.25 . 10^{-2}$	$0.086 . 10^{-2}$	2·9	$6.7 . 10^{-6}$	70
Second subject (D.L.M.)	0·31	0·165	1·88	16·1	44

In Wright's investigation (1941), the subjects determined equal colour steps on various lines in the C.I.E. chart, and in figure 15 the logarithm of the observed colour step is plotted against distance from one end of the line for four of the principal lines studied. Wright's colour steps were not actual limens but
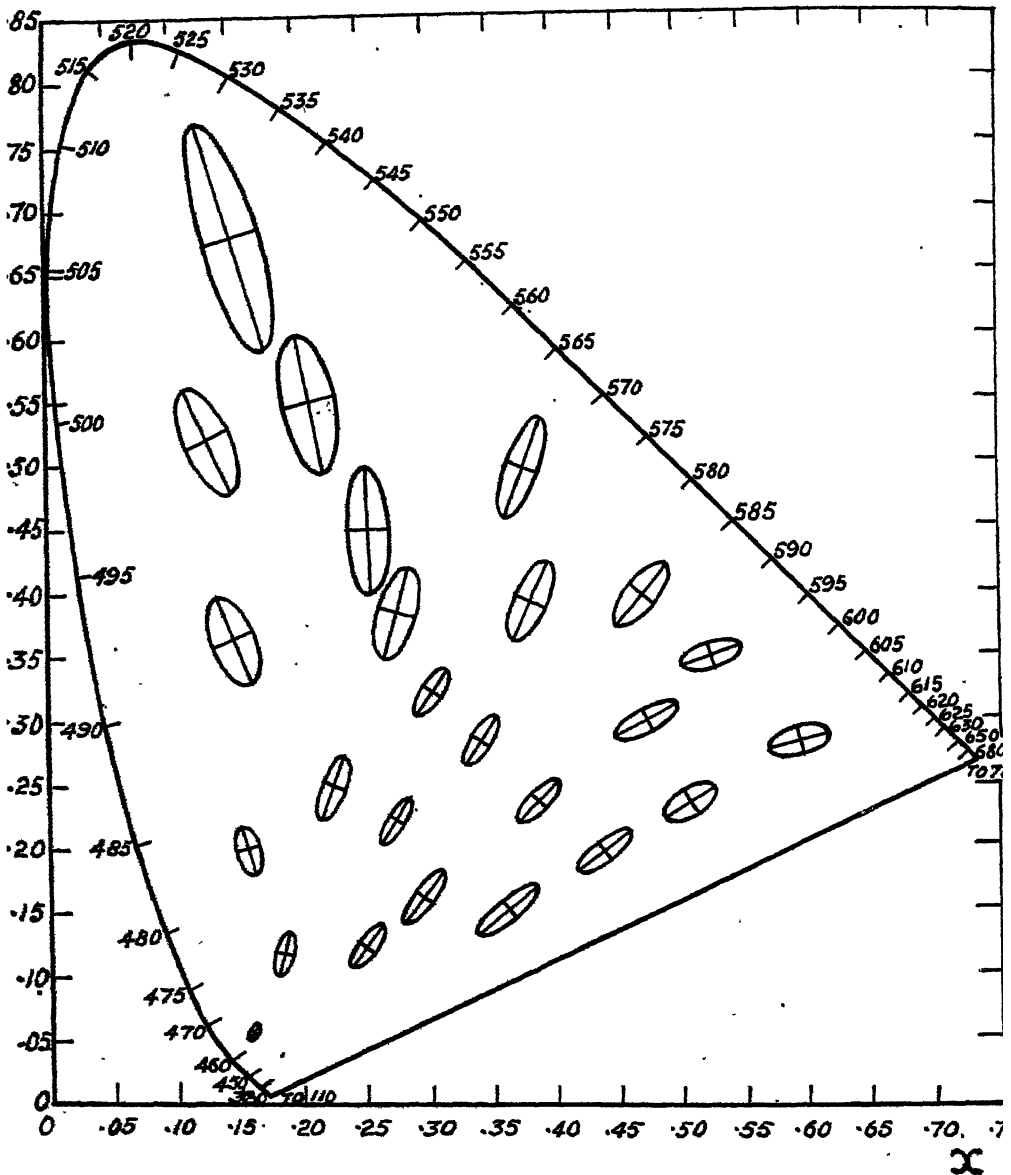
Figure 10.   MacAdam's " standard deviation " ellipses for P.G.N.

. " equal small colour differences " ; he found that after practice a subject could maintain a reasonably steady criterion.   On the average, the colour step was about three times the limen.   In reporting his results, Wright reduced the steps for the four subjects to about the same average values by applying suitable constant factors : figure 15 shows the results before the application of these factors.

No direct calculations from the line-element were made for comparison with Wright's results, but a fair idea of the variations along the several lines was obtained by interpolation from the 25 computed ellipses of figure 9. A similar derivation was made from MacAdam's ellipses for P.G.N (figure 10). In figure 15,
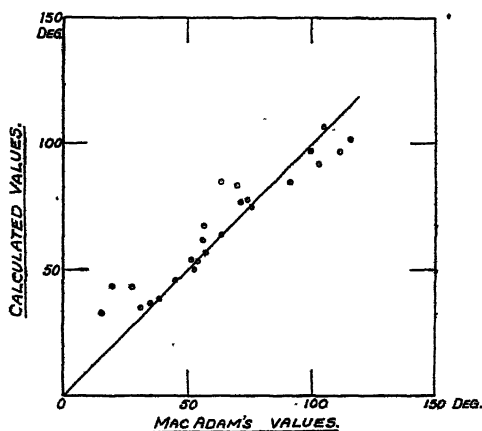
Figure 11. Inclination of the major axis of ellipse to the axis of *x*.
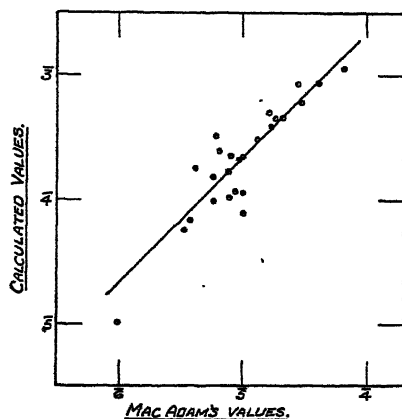
Figure 12. Log (area of ellipse).

the computed colour limens times 3 are shown as the thick-line graphs and MacAdam's standard deviations times 9 as the broken-line graphs. These factors allow for the average systematic differences between small colour step and limen, limen and standard deviation, as estimated by Wright and MacAdam
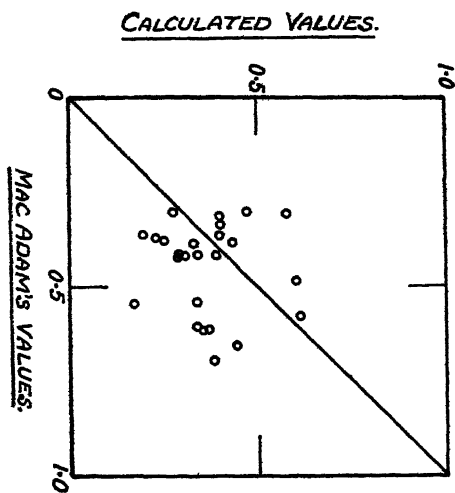
Figure 13. Log $\left(\dfrac{\text{major axis of ellipse}}{\text{minor axis of ellipse}}\right)$.
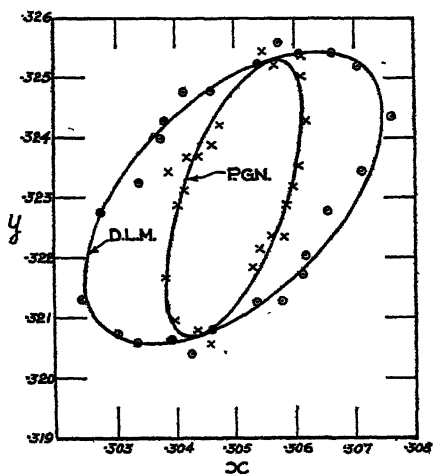
Figure 14. G.C.L. ellipses for two subjects. (MacAdam).

respectively. Admitting that these factors may differ for different subjects, and that F may also differ, it is fair to suppose the four line-element curves in figure 15 all shifted up or down by the same amount when comparing with the experimental curves of each subject. Even then there is no close agreement with
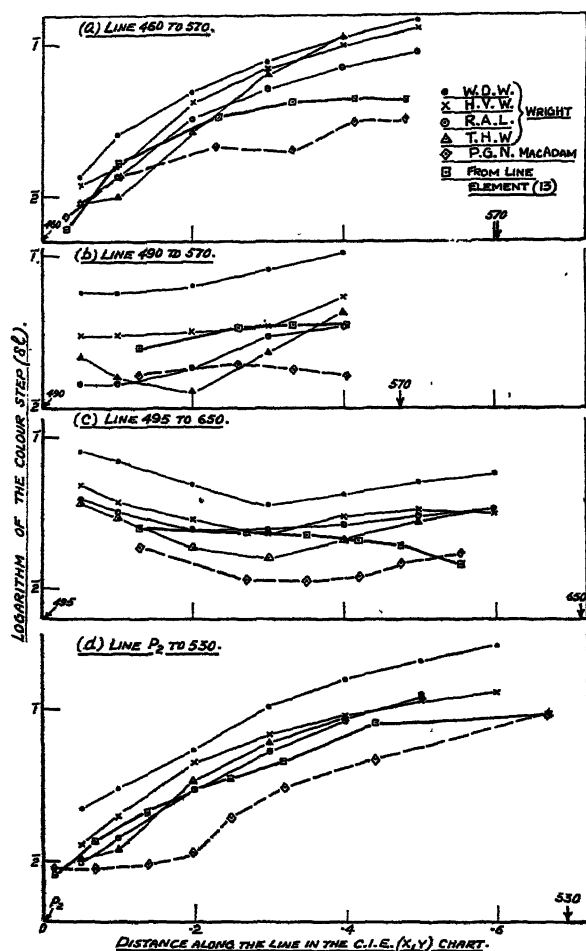
Figure 15.

the curves of any of the five subjects. It may be noted that the experimental curves of different subjects, compared in the same way, show differences which are hardly less marked.

§ 10. DISCUSSION

The line-element may be regarded as a mathematical model of the mechanism underlying a particular group of visual properties (increment threshold, colour-matching relations, visibility curve, etc.). In the Helmholtz and modified elements, the respective terms correspond to three component mechanisms, with each of which is associated a unique spectral sensitivity curve. A difference between adjacent light patches will be perceived by the action of one of these component mechanisms, say the red, if the product,

$$|\delta x|\left(\frac{\sqrt{3}}{a+x}\right)\text{(Helmholtz element)}\quad\text{or}\quad |\delta R|\left(\frac{\zeta(R)}{\rho}\right)\text{ (modified element)},$$

exceeds F. The first factor in these products is the difference of stimulation of the red mechanism in the two patches, evaluated on the basis of the red spectral-sensitivity curve. The second factor defines the sensitivity of the red mechanism

in the state to which it is brought by the total stimulation, x or R, which may be taken here as approximately the same in the two patches. The product, on this view, is a measure of the response of the red mechanism to the difference of stimulation. The only interaction between the red, green and blue component mechanisms contemplated by these line-elements is a certain degree of summation of their responses, $\dfrac{\delta x \sqrt{3}}{a+x}$, etc., according to the sums of squares relation.

The Schrödinger element (3) is an example of the wider class represented by

$$\delta s^2 \equiv [\delta x \cdot g_x(x, y, z)]^2 + [\delta y \cdot g_y(x, y, z)]^2 + [\delta z \cdot g_z(x, y, z)]^2 = F^2. \qquad \ldots \ldots (28)$$

For line-elements in this class, the component mechanisms preserve their individuality—with each is associated a unique spectral sensitivity curve *—but they are so linked that the state of each as indicated by the sensitivity factors g_x, g_y, g_z depends on the stimulations of all three mechanisms. This linkage is additional to the summation of responses according to the sum of squares relation.

For the above elements, the axes of the small ellipsoid defined by the line-element at each point of f.c. space are parallel to the co-ordinated axes. In the more general element,

$$\delta s^2 \equiv \sum_{m,\,n} g_{mn} \delta x_m \delta x_n = F(m, n = 1, 2, 3), \qquad \ldots \ldots (29)$$

the axes of the elementary ellipsoids at different points are not necessarily parallel to the axes of the ellipsoid at any one point in the space, but different rotations are in general required for different points. A rotation of the axes corresponds to a change of the primaries so that on this model the spectral-sensitivity curves of the three mechanisms are progressively modified as the stimulation (defined by the position in b.c. space) is varied.

To what extent does the simple model defined by the modified line-element (13) fit the facts? The matter is complicated by the very considerable differences in the experimental results of different subjects. To be satisfactory, the line-element (13) with a particular numerical determination of R_λ, G_λ, B_λ, ρ, γ, β, F and $\zeta(p)$ should reproduce the colour-matching values, the step-by-step visibility curve, the increment and colour limens and possibly some other visual properties of a particular eye, or, more precisely, of a particular rod-free area of a given retina. Changes in the visual properties in passing from one rod-free area to another in the same or a different eye should all be accounted for at a stroke by a change in the numerical determination of the line-element. For the comparisons made in this paper, a single numerical determination has been used : for the distribution coefficients $\bar{x}, \bar{y}, \bar{z}$ the C.I.E. values, for the fundamental response curves the function defined in (15) for the limiting Fechner fractions numbers in the proportion $0 \cdot 78 : 1 : 4 \cdot 46$, and for F the value $0 \cdot 01$. While certain main features of the experimental results are reproduced, satisfactory quantitative agreement with the measurements of all the visual properties for some particular subject has not been demonstrated. In fact no such complete set of measurements is available. For the restricted group of

---

* Except in the degenerate case when any two of g_x, g_y, g_z stand in a constant ratio.

colour-limen measurements, the comparisons made (figures 8 to 15) show differences which could no doubt be reduced by a different numerical determination of the line-element for each subject. But in the case of MacAdam's subject P.G.N. it can be shown that complete agreement could not be reached by any change in which the fundamental response functions R_λ, G_λ, B_λ remained always positive, linear forms in the C.I.E. distribution coefficients. (See Appendix.) Our conclusion must be that the modified line-element, constructed to suit measurements of increment limens, leads to the right kind of step-by-step visibility curve, but is in difficulties when applied to measurements of colour limens.

§ 11. ACKNOWLEDGMENT

This paper is published by permission of the Director of the National Physical Laboratory.

APPENDIX

The Gaussian curvature of the surface of constant C.I.E. brightness according to the modified line-element

Following Silberstein (1943), the unit C.I.E. co-ordinates x, y for colours of a given constant C.I.E. brightness can be regarded as curvilinear co-ordinates on a surface embedded in a three-dimensional Euclidean space, the distance apart of neighbouring points on this surface being defined by the expression for the line-element δs. This expression, for the modified line-element, is the square root of the right-hand side of equation (27). Silberstein uses an empirical expression for δs derived from MacAdam's measurements for the subject P.G.N. (figure 10), and shows that the Gaussian curvature K of the surface is not only not constant, but that it is positive in some areas, negative in others.

The Gaussian curvature is invariant to transformations, linear or non-linear, of the co-ordinates x, y, and it is permissible to employ R, G in place of x, y. For the modified line-element, an expression for K is obtained quite easily if we accept for $\zeta(p)$ the approximation $\dfrac{9}{1+9p}$, to which it reduces when p exceeds 1. This corresponds to using curve $\eta(p)$ instead of curve $\xi(p)$ (figure 2). K is then given by

$$K = \frac{\rho^2\gamma^2\beta^2(1+9R)(1+9G)(1+9B)v_rv_gv_b(9y' + v_r + v_g + v_b)}{[\rho^2v_r{}^2(1+9R)^2 + \gamma^2v_g{}^2(1+9G)^2 + \beta^2v_b{}^2(1+9B)^2]^2}, \quad \ldots\ldots(30)$$

where $y' = v_rR + v_gG + v_bB = \text{const.}$ This condition is obtained by solving the equation (25) for x', y', z'; it expresses the fact that the colours considered all have a constant C.I.E. brightness (equal to $2 \cdot 19 . 10^5 y'$ photons).

Since for real colours R, G and B are all positive, it is clear that the Gaussian curvature will have the same sign at all points of the surface of constant C.I.E. brightness.

Computation gives

$$v_r = \quad 2 \cdot 70 \times 10^{-4}$$
$$v_g = \quad 4 \cdot 07 \times 10^{-4}$$
$$v_b = -4 \cdot 76 \times 10^{-5}$$
$$v_r + v_g + v_b = \quad 6 \cdot 29 \times 10^{-4}.$$

Thus $v_r v_g v_b (9y' + v_r + v_g + v_b)$, and hence K, is negative for any value of y', i.e. for any C.I.E. brightness.

It follows that the modified line-element is inconsistent with MacAdam's results for P.G.N., which for different colours yield K values of opposite sign. This inconsistency persists however ρ, γ, β be chosen and whatever change be made in the fundamental response curves R_λ, G_λ, B_λ provided these remain linear forms in the C.I.E. distribution coefficients, and lead to all positive co-ordinates R, G, B for real colours.

REFERENCES

COBLENZ, W. W. and EMERSON, W. B., 1917.   *Sci. Pap. U.S. Bur. Stand.* No. 303.
GIBSON, K. S. and TYNDALL, E. P. T., 1923.   *Sci. Pap. U.S. Bur. Stand.* No. 475.
HECHT, S., 1932.   *Discussion on Vision*, Phys. and Opt. Societies, 126.
HELMHOLTZ, H. VON, 1891.   *Z. Sinnesphysiol.* 3, 517.
KÖNIG, A. and DIETERICI, C., 1884.   *Ann. Phys., Lpz.*, 22, 579.
MACADAM, D. L., 1942.   *J. Opt. Soc. Amer.* 32, 247.
PRIEST, I. G. and BRICKWEDDE, F. G., 1926.   *J. Opt. Soc. Amer.* 13, 306.
SCHRÖDINGER, E., 1920.   *Ann. Phys., Lpz.*, 63.
SILBERSTEIN, L., 1943.   *J. Opt. Soc. Amer.* 33, 1.
STILES, W. S., 1939.   *Proc. Roy. Soc.* B, 127, 64.
WALTERS, H. V. and WRIGHT, W. D., 1943.   *Proc. Roy. Soc.* B, 131, 340.
WRIGHT, W. D., 1936.   *J. Physiol.* 87, 23.
WRIGHT, W. D., 1941.   *Proc. Phys. Soc.* 53, 93.
WRIGHT, W. D. and PITT, F. H. G., 1934.   *Proc. Phys. Soc.* 46, 459.

ON DECENTRED ASPHERIC PLATES

By E. H. LINFOOT,

University of Bristol

ABSTRACT.   The main purpose of the paper is to discuss the effects of errors of centring on the optical performance of plate-mirror systems.   Such errors may occur during the aspheric grinding of the plates and also in the lining up of the system.   It is first shown that if a grinding technique is used which builds up the asphericity at a proportionately equal rate all over the surface, the result of a large number of small centring errors at different stages of the grinding is very nearly equivalent to a simple decentring of the asphericity on the surface, combined with a small amount of primary astigmatism.   Next, general formulae are obtained for the effects on the Seidel errors of a centred system of decentring and tilting its components, and it is shown that in the case of a plate-mirror system the formulae can be used to estimate the practical tolerances for disturbances of this kind.   Lastly, the general formulae are applied to discuss the lining up of a two-sphere one-plate Schmidt-Cassegrain camera.

§1. EFFECT OF ERRORS IN GRINDING

IN the construction and use of plate-mirror systems, problems of centring are met with at two stages.   The first is during the actual aspherizing of the surfaces ; the second during the lining up of the system.   An error in centring at the first stage will in general lead to an asymmetrical error in the figure of the surface.   This will result in a slight change in the centring chosen

for the surface when the complete system comes to be lined up, for the centring chosen will be that at which the optical effect of the asymmetry appears as small as possible. Thus the practical effect of the asymmetry depends not so much on its actual amount as on the extent to which it cannot be compensated by lateral shift of the surface which bears it.

Broadly speaking, the most successful methods of figuring aspheric plates are those in which the asphericity is built up at a proportionately equal rate all over the surface. One such method, well suited to the resources of the amateur astronomer, has been described in an ably written paper by H. W. Cox (1939). It depends on the use of a flexible grinding lap consisting of lead facets arranged in petal-shaped areas on a Sorbo rubber backing, operated by a mechanical traverse. The lap is traversed in short, fairly rapid strokes as the plate slowly rotates. With such an arrangement, asymmetry arising from random variations in the stroke is eliminated for all practical purposes, and inaccuracy of centring of the plate in its holder, or of the plate-holder on its spindle, becomes a potentially important source of error.

To estimate the practical effect of errors arising in this way, we observe that if the grinder is adjusted to generate a true curve, with only small relative errors, whose depth increases at proportionately equal rates all over the surface, then the effect of having the plate wrongly centred for part of the time is to contribute a weak " Schmidt plate " with displaced centre to the figure instead of adding its strength to the main (i.e. truly centred) figuring.

Suppose that m errors of this type are made and that the plate is centred by trial in the final assembly. The plate figure is then the sum of m figurings, each with a slight decentring, whose strengths together make up the full plate-strength B. We can investigate the properties of such a plate by applying the methods of Conrady (1919).

The deviation (Δx, Δy) of the point of intersection with the Petzval surface of the ray through the point P with polar co-ordinates (S, E) in the entry pupil is given in the case of pure primary spherical aberration by

$$\Delta x = BS^3 \cos E, \qquad \Delta y = BS^3 \sin E,$$

where the coefficient B measures the amount of the aberration. From a decentred spherical aberration of centre (ξ_r, η_r) and amount B_r, we therefore have deviation-contributions

$$\Delta x = B_r S_r^3 \cos E_r, \qquad \Delta y = B_r S_r^3 \sin E_r,$$

where (see figure 1)

$$S_r \cos E_r = S \cos E - \xi_r, \qquad S_r \sin E_r = S \sin E - \eta_r ;$$

$$S_r^2 = S^2 - 2S(\xi_r \cos E + \eta_r \sin E) + (\xi_r^2 + \eta_r^2).$$

If ξ_r, η_r are so small that their squares can be neglected, the deviation-contributions can be written

$$\Delta x = B_r [S^2 + (\xi_r^2 + \eta_r^2) - 2S(\xi_r \cos E + \eta_r \sin E)](S \cos E - \xi_r)$$

$$= B_r S^3 \cos E - B_r \xi_r S^2 (2 + \cos 2E) - B_r \eta_r S^2 \sin 2E,$$

$$\Delta y = B_r S^3 \sin E - B_r \xi_r S^2 \sin 2E - B_r \eta_r S^2 (2 - \cos 2E),$$

and the total deviations effected by the plate are accordingly

$$\Delta x = \left(\sum_{r=1}^{m} B_r \right) S^3 \cos E - \left(\sum_{r=1}^{m} B_r \xi_r \right) S^2 (2 + \cos 2E) - \left(\sum_{r=1}^{m} B_r \eta_r \right) S^2 \sin 2E,$$

$$\Delta y = \left(\sum_{r=1}^{m} B_r \right) S^3 \sin E - \left(\sum_{r=1}^{m} B_r \xi_r \right) S^2 \sin 2E - \left(\sum_{r=1}^{m} B_r \eta_r \right) S^2 (2 - \cos 2E),$$

$$\ldots\ldots(1)$$

i. e. are those corresponding to primary spherical aberration of amount $B = \sum_{r=1}^{m} B_r$, together with a coma contribution which is measured by the sum of the vectors $B_r(\xi_r, \eta_r)$. This coma will be annulled if the plate is given a lateral displacement (Ξ, H), where

$$B\Xi = - \sum_{r=1}^{m} B_r \xi_r, \qquad BH = - \sum_{r=1}^{m} B_r \eta_r, \qquad \ldots\ldots(2)$$

when the system is assembled. Since the plate will in practice be centred by trial, such a displacement will be given automatically. The largest errors then
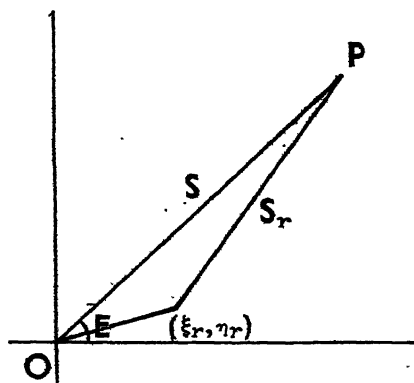
Figure 1.

remaining are the second-order terms (neglected above), and, as will appear later, these give rise to astigmatism, in second-order small quantity, uniform over the image-field.

In certain important cases * this astigmatism also can be compensated by adjustments of the components perpendicular to the axis ; it follows that in these cases a first-order centring error in aspherical grinding, *provided it is of the type specified above*, leads only to errors of the third order of small quantities in the finally adjusted system, i.e. to extra terms in the expressions for the displacements $\Delta x, \Delta y$ in the final system whose coefficients are third-order small quantities. Such terms have no perceptible effect on the physical image in many practical cases. In general, a second-order amount of astigmatism remains after the coma has been removed by empirical plate-adjustment.

The general conclusion is that, provided the grinding is arranged so as to generate the asphericity at proportionately equal rates all over the surface, the

---

* For example, in the monocentric Schmidt-Cassegrain systems and in the two-plate Schmidt-Cassegrains.

effect of the centring errors, which must always be present in some degree, is far less serious than might appear at first sight. It seems likely that this circumstance plays a large part in determining the success or failure of a grinding technique.

§ 2. EMPIRICAL LINING-UP OF A CENTRED OPTICAL SYSTEM

Consider the "equivalent plate system".* Let $B_1, B_2 \ldots B_n$ be the plate-strengths ; $\sigma_1, \sigma_2 \ldots \sigma_n$ the respective distances of the (real or virtual) plates in front of the entry pupil ; $(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots (\xi_n, \eta_n)$ their respective lateral displacements. For the present, we suppose the components of the system to be free from errors of figure.

Then, as before, the displacement-contribution from the first plate to the intersection with the Petzval surface of a ray, originally parallel to the axis, through the point of the entry pupil whose polar co-ordinates are (S, E), is

$$\Delta x = B_1 S_1^3 \cos E_1, \qquad \Delta y = B_1 S_1^3 \sin E_1,$$

where

$$S_1 \cos E_1 = S \cos E - \xi_1, \qquad S_1 \sin E_1 = S \sin E - \eta_1,$$

$$S_1^2 = S^2 + (\xi_1^2 + \eta_1^2) - 2S(\xi_1 \cos E + \eta_1 \sin E).$$

Thus

$$\Delta x = B_1[S^2 + (\xi_1^2 + \eta_1^2) - 2S(\xi_1 \cos E + \eta_1 \sin E)][S \cos E - \xi_1]$$
$$= B_1[S^3 \cos E - S^2(\xi_1(2 + \cos 2E) + \eta_1 \sin 2E)$$
$$\qquad + S((3\xi_1^2 + \eta_1^2) \cos E + 2\xi_1 \eta_1 \sin E) - \xi_1(\xi_1^2 + \eta_1^2)],$$

$$\Delta y = B_1[S^3 \sin E - S^2(\xi_1 \sin 2E + \eta_1(2 - \cos 2E))$$
$$\qquad + S((\xi_1^2 + 3\eta_1^2) \sin E + 2\xi_1 \eta_1 \cos E) - \eta_1(\xi_1^2 + \eta_1^2)].$$

If we write

$$\xi_1 + i\eta_1 = \rho_1 e^{i\phi_1},$$

these equations can be given the more compact form

$$\Delta x + i\Delta y = B_1 e^{i\phi_1}[S^3 e^{i(E - \phi_1)} - S^2 \rho_1 (2 + e^{2i(E - \phi_1)})$$
$$\qquad + S\rho_1^2 e^{i(E - \phi_1)}(2 + e^{2i(E - \phi_1)}) - \rho_1^3].$$

For the complete plate-system we therefore have, in Seidel approximation,

$$\Delta x + i\Delta y = S^3 e^{iE} \sum_{r=1}^{n} B_r - S^2 \left(2 \sum_{r=1}^{n} B_r \rho_r e^{i\phi_r} + e^{2iE} \sum_{r=1}^{n} B_r \rho_r e^{-i\phi_r} \right)$$

$$+ S \left(2 e^{iE} \sum_{r=1}^{n} B_r \rho_r^2 + e^{-iE} \sum_{r=1}^{n} B_r \rho_r^2 e^{2i\phi_r} \right) - \sum_{r=1}^{n} B_r \rho_r^3 e^{i\phi_r}, \quad \ldots\ldots(3)$$

where B_r denotes the strength of the rth plate and $\rho_r e^{i\phi_r} = \xi_r + i\eta_r$, its lateral displacement. Here Δx, Δy measure the lateral displacements in the Petzval surface and $B_1, B_2, \ldots\ldots$ measure the plate-strengths, both in suitable units. The connection between the two sets of units can be expressed by saying that if for the unit of plate-strength we choose the strength of the Schmidt plate of a concave spherical mirror of paraxial focal length f_1, then the actual linear displacements δx, δy in the Petzval surface are obtained on multiplying Δx, Δy by $f/8f_1^3$, where f is the focal length of the system.

* In the sense of the method of plate-diagram analysis. *See* C. R. Burch, *Mon. Not, R. Astr. Soc.* **102**, 159–165 (1942) ; E. H. Linfoot, *ibid.* **104**, 48–64 (1944), for an account of this method.

The first three terms on the right-hand side of equation (3) are easily recognized, on separating them into real and imaginary parts once more, as arising respectively from spherical aberration, coma, and astigmatism combined with change of focus. The fourth term, being independent of E and S, represents a lateral shift of the image as a whole.

We shall suppose the system designed so that $\sum\limits_{r=1}^{n} B_r = 0$, i.e. spherically corrected. Then the conditions for the axial image to be error-free are, if we do not permit a change of focus,

$$\sum_{r=1}^{n} B_r \rho_r e^{i\phi_r} = 0, \qquad \sum_{r=1}^{n} B_r \rho_r^2 e^{2i\phi_r} = 0, \qquad \sum_{r=1}^{n} B_r \rho_r^2 = 0. \quad \ldots \ldots (4)$$

In practice, however, the focus is usually determined empirically and the term $2S e^{iB} \sum\limits_{r} B_r \rho_r^2$ will not show itself as a defect in the image. It follows that the conditions to be satisfied are

$$\sum_{r=1}^{n} B_r = 0, \qquad \sum_{r=1}^{n} B_r \rho_r e^{i\phi_r} = 0, \qquad \sum_{r=1}^{n} B_r \rho_r^2 e^{2i\phi_r} = 0. \quad \ldots \ldots (5)$$

We can express this result as follows : *If the (real or virtual) elements of the plate-system associated with an optical system are given lateral displacements $v_1, v_2 \ldots$ (v a vector) satisfying the conditions*

$$\sum_{r} B_r v_r = 0, \qquad \sum_{r} B_r v_r^2 = 0, \qquad\qquad \ldots \ldots (6)$$

the axial image will not develop spherical aberration, coma or astigmatism. It will be laterally displaced by an amount proportional to $\sum\limits_{r} B_r |v_r|^2 v_r$, and will be moved along the axis, away from the system, through a distance proportional to $2\sum\limits_{r} B_r |v_r^2|$.

Equation (3) gives a simple proof of the property, mentioned in §1, that when a plate suffering from errors of figure of the type there described is centred empirically, the residual error is effectively primary astigmatism, in second-order small quantity. It is only necessary to set $\sigma_1 = \sigma_2 \ldots \sigma_m$ in the present discussion ($\sigma_{m+1}, \ldots \sigma_n$ remaining arbitrary) and to observe that the operation of centring the plate empirically is expressed by adding to each of $\rho_1 e^{i\phi_1}, \ldots \rho_m e^{i\phi_m}$ the complex constant

$$\alpha = - \sum_{r=1}^{n} B_r \rho_r e^{i\phi_r} \Big/ \sum_{r=1}^{m} B_r.$$

After this operation, the coefficient of S^2 in (3) vanishes and the third term becomes the leading term.

Off-axis images

To determine the effect on the off-axis errors of a system of decentring its plates, we consider a parallel beam coming in at an angle V with the axis, the azimuth of this angle being χ. The principal ray meets a plane distant σ in fron of the entry pupil in the point

$$X = \sigma V \cos \chi, \qquad Y = \sigma V \sin \chi,$$

and the effect of the plate is, therefore, the same (in Seidel approximation) as if it were situated in the entry-pupil but with a lateral displacement $(-\sigma V \cos \chi, -\sigma V \sin \chi)$, and the pencil were parallel to the axis.* It follows that the off-axis errors for laterally displaced plates are obtained on replacing (ξ_r, η_r) by $(\xi_r - \sigma_r V \cos \chi, \eta_r - \sigma_r V \sin \chi)$ in (3). Set

$$\xi_r - \sigma_r V \cos \chi = P_r \cos \Phi_r, \qquad \eta_r - \sigma_r V \sin \chi = P_r \sin \Phi_r,$$

which gives

$$P_r e^{\pm i \Phi_r} = \rho_r e^{\pm i \phi_r} - \sigma_r V e^{\pm i \chi},$$

and we obtain in place of (3) the equation

$$\Delta x + i \Delta y = S^3 e^{iE} \sum_{r=1}^{n} B_r - S^2 \left(2 \sum_{r=1}^{n} B_r P_r e^{i \Phi_r} + e^{2iE} \sum_{r=1}^{n} B_r P_r e^{-i \Phi_r} \right)$$

$$+ S \left(2 e^{iE} \sum_{r=1}^{n} B_r P_r^2 + e^{-iE} \sum_{r=1}^{n} B_r P_r^2 e^{2i\Phi_r} \right) - \sum_{r=1}^{n} B_r P_r^3 e^{i\Phi_r}. \quad \ldots \ldots (7)$$

The four terms of this new equation have the same significance as before, and they now specify the spherical aberration, coma, astigmatism with change of focus, and lateral displacement † at any point of the image-field of the system with decentred plates.

We assumed above that the optical components were free from errors of figure. The same analysis suffices, however, to cover the more general case where the plates suffer from errors of figure of the type discussed in § 1. It is only necessary to represent each plate as a cluster of mutually decentred, error-free component-plates held rigidly in contact, and to apply (7) to the system of component-plates so obtained.

An inspection of the terms on the right of (7) enables us to draw one or two general conclusions about the effect of these errors of figure on the performance of an empirically lined up system.

Since the first term $S^3 e^{iE} \sum_r B_r$, which represents the spherical aberration, does not involve V or the quantities ρ_r, ϕ_r, it follows that the spherical aberration of the system is unaffected by the errors and remains constant all over the field.

The second term, which represents the coma, falls into two parts :

$$- S^2 (2 \Sigma_r B_r e^{i\Phi_r} + e^{2iE} \Sigma_r B_r P_r e^{-i\Phi_r})$$

$$= - S^2 (2 \Sigma_r B_r \rho_r e^{i\phi_r} + e^{2iE} \Sigma_r B_r \rho_r e^{i\phi_r})$$

$$+ S^2 V (2 e^{i\chi} + e^{2iE - i\chi}) \Sigma_r B_r \sigma_r. \quad \ldots \ldots (8)$$

The second part represents off-axis coma of the usual kind, and vanishes identically if, as we suppose, the system is designed to satisfy the aplanatism condition $\Sigma_r B_r \sigma_r = 0$. The first part does not depend on V or χ, and so represents a coma uniform over the whole field, which can be eliminated by a lateral shift of any one

* The effect of the small relative tilt between plate and off-axis pencil is zero in Seidel approximation.

† The last term is associated with the distortion of the system, but does not by itself measure the distortion. See *Mon. Not. R. Astr. Soc.* **104**, 49 (1944).

of the plates in the plate-system, or of several plates together, so as to secure the fulfilment of the condition $\Sigma_r B_r P_r e^{i\Phi_r} = 0$.

The third term,

$$S(2e^{iE}\Sigma_r B_r P_r^2 + e^{-iE}\Sigma_r B_r P_r^2 e^{2i\Phi_r})$$

$$= S(2e^{iE}\Sigma_r B_r \rho_r^2 + e^{-iE}\Sigma_r B_r \rho_r^2 e^{2i\phi_r})$$

$$- 2SV(2e^{iE}\Sigma_r B_r \rho_r \sigma_r \cos(\phi_r - \chi) + e^{-iE}\Sigma_r B_r \rho_r \sigma_r e^{i(\phi_r + \chi)})$$

$$+ SV^2(2e^{iE} + e^{2i\chi - iE})\Sigma_r B_r \sigma_r^2, \qquad \dots \dots (9)$$

represents astigmatism and focus shift.* It falls into three parts, in accordance with (9) ; the third part contains the ordinary off-axis astigmatism of the system, which vanishes in systems (the so-called Seidel anastigmats) in which $\Sigma_r B_r \sigma_r^2 = 0$. The remaining parts represent the effect of the decentring ; we shall consider them in more detail, in the next section, in the particular case of a three-plate system.

Finally, the fourth term,

$$\Sigma_r B_r P_r^3 e^{i\phi_r} = \Sigma_r B_r P_r e^{i\phi_r} |P_r e^{i\phi_r}|^2$$

$$= \Sigma_r B_r \rho_r^3 e^{i\phi_r} - V\Sigma_r B_r \rho_r^2 \sigma_r(2e^{i\chi} + e^{2i\phi_r - i\chi_r})$$

$$+ V^2\Sigma_r B_r \rho_r \sigma_r^2(2e^{i\phi_r} + e^{2i\chi - i\phi_r})$$

$$- V^3 e^{i\chi}\Sigma_r B_r \sigma_r^3, \qquad \dots \dots (10)$$

which represents the plate-contribution to the distortion of the system, remains independent of S and E when the plates are decentred, and so does not affect the sharpness of the images. But it does depend on V and on the quantities ρ_r, ϕ_r, so that, as was pointed out by Conrady in the paper already referred to, a system lined up to give satisfactory images may still contain centring errors which appreciably modify its distortion.

The above analysis applies in principle to ordinary lens systems with spherical surfaces ; their associated plate systems, to which we can apply (7), consist entirely of virtual plates. But the application is not very satisfactory in the case of wide-aperture lens systems, since in these the Seidel theory is far from giving a true picture of the image-errors, especially in the outer parts of the field. In its application to plate-mirror systems of the Schmidt-Cassegrain type, however, the theory supplies information of practical value even at nominal apertures as high as F/2. This different state of affairs can be attributed to the much smaller size of the fifth-order error-coefficients which results when mirrors are used in place of lenses as optical power-surfaces, together with the circumstance that vignetting restricts the field diameter of an F/2 Schmidt-Cassegrain to something less than 15°.

* An incidental advantage of the complex notation used in this section is that the terms representing focus-shift appear separated from those representing astigmatism. The former are the terms in Se^{iE}, the latter those in Se^{-iE}.

§3. APPLICATION TO SCHMIDT-CASSEGRAIN CAMERAS

As a practical example, we consider the lining-up of a Schmidt-Cassegrain camera consisting of two spherical mirrors and one aspheric plate. The general appearance of such a system is shown in figure 2.

The optical lining-up of a correctly figured system of this type can be carried out as follows :

(*a*) place the primary mirror M_1 symmetrically in its mounting by measurement ;

(*b*) place the plate P at the measured distance from M_1 and centre it by measurement ;

(*c*) move the secondary mirror M_2 along the axis, keeping it roughly centred, until spherical aberration is zero, when P is moved laterally so as to annul coma, and thus allow the spherical aberration to be observed ;

(*d*) *tilt* M_2, readjusting P laterally each time, until the astigmatism is zero on axis ;

(*e*) move P along the axis until the coma remains zero for small image-displacements off axis.
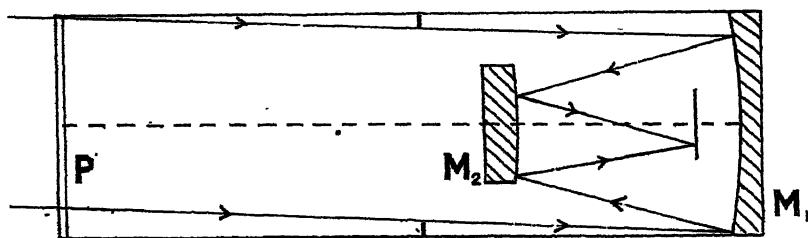
Figure 2.

The system is then in " Seidel adjustment ". This adjustment must now be altered, by an amount which depends on the field to be covered, in order to balance the higher errors by suitable amounts of opposing Seidel error over the required field.

In the monocentric Schmidt-Cassegrains, the change consists in moving M_2 and P along the axis to balance the lateral spherical aberration against primary spherical aberration, while keeping the coma zero. This gives a system which is extremely close to a monocentric system. The last operation is always to adjust P laterally to bring the coma to zero on axis. In the flat-fielded aplanats, the compromises to be made are of a more complicated nature.

In a practical case, however, the plate-figure usually contains a little primary astigmatism arising (in the manner already described) from centring errors during grinding. The amount of this astigmatism can be kept within tolerable limits by the use of proper precautions during the aspherizing process. For example, the plate of an F/2 Schmidt-Cassegrain system might be regarded as satisfactory if the astigmatic confusion-circle corresponding to its astigmatism of figure did not exceed $1/5000$ inch in diameter. But the presence of this astigmatism complicates the task of lining up the system ; if the above procedure be adopted it results in a decentred line-up which Seidel theory and experiment agree in condemning as unsatisfactory in most cases because of its effect on the

images in the outer parts of the field.* It is therefore better to begin by placing the two mirrors M_1, M_2 in correct alignment by mechanical means (with provision for varying the position of M_2 along the axis) and to confine the empirical adjustments to operations (*c*), (*e*) and the annulling of coma by lateral plate-shift.

The question then arises, how accurate the mechanical alignment of M_1 and M_2 needs to be. We shall investigate this by applying the general formulae of the last section. To be of practical value, the investigation must take account of the higher aberrations as well as the Seidel properties of the system. In the case of wide-aperture lens systems, where the higher aberrations have to be carefully compensated over a relatively large field, a general theoretical discussion of centring tolerances would seem to be hardly practicable at the present time, since it would at once involve us in the formidable analytical complications of fifth-order optics. We should need to rely instead upon conclusions drawn from experiment, and these conclusions would only remain valid for systems very close in design to those on which the experimental observations had been made. In the case of the Schmidt-Cassegrain systems, comprising only three or four effective optical surfaces and working over a smaller field, high definition is a consequence of the intrinsic smallness of the higher aberrations of each element rather than of their mutual compensation.† A small error in the lining up of the system therefore results in only a small percentage change in the amounts of its higher aberrations. Since the Seidel coma and the primary spherical aberration are under continuous control,‡ it is the size of these higher aberrations which determines the actual confusion-spreads in a Seidel anastigmat (or near-anastigmat) in correct adjustment. It follows that small centring errors will not markedly affect the confusion-spreads of a Schmidt-Cassegrain system provided that their contribution to the Seidel errors is small compared with the image-spreads of the correctly adjusted system. A more convenient formulation, practically equivalent to the first, is that the centring errors can be tolerated if their Seidel contribution to the displacement $\Delta x + i\Delta y$ is everywhere small compared with the allowed upper limit to the size of this quantity for the system.

We can obtain a reasonable estimate of the amount of decentring which can be tolerated in the line-up of the secondary mirror of a Schmidt-Cassegrain camera by laying down that the resulting final disturbance $\delta x + i\delta y$ of the Seidel displacement-vector in the Petzval surface must nowhere exceed one fifth of the allowed image-spread for the system. In a practical case, the allowed image-spread might be 0·001 inch, and the tolerance would then be given by the condition

$$|\,\delta x + i\delta y\,| < \frac{1}{5000}\,\text{inch}$$

over the whole field.

---

* The monocentric Schmidt-Cassegrains are to a certain extent an exception to this rule.

† Higher spherical aberration (zonal aberration) appears as an exception to this statement, inasmuch as it is usually fully compensated in the plate-figuring. But this aberration also is intrinsically very small, as can be shown in a striking manner in a particular case by removing the secondary mirror of an F/2 two-sphere aplanat, placing the plate in contact with the primary and Foucault-testing the pair with a pinhole at an adjustable distance along the axis. When the pinhole-distance is suitably chosen, residual spherical aberration is found to be scarcely detectable with the knife-edge.

‡ They can be varied at will by altering the spacings of the components.

The disturbance of the Seidel properties due to centring errors in the line-up can be calculated from the general formulae of the last section as soon as the errors have been expressed in terms of displacements of the elements in the associated plate system. To do this, we image the surface-figurings, aspheric plates and missing Schmidt plates of the system through into object-space in the usual way, taking account of the errors in tilt and in lateral adjustment of the mirrors when calculating the images according to Gaussian optics. The associated plate system thus obtained has its elements laterally displaced and tilted. The tilts of the elements have no effect on the Seidel aberrations and are therefore disregarded ; their lateral displacements are a set of parameters which describe, sufficiently fully for the present purpose, the errors in the line-up of the original system.

A Schmidt-Cassegrain with both mirrors spherical is a " three plate system " ; that is to say, its plate-diagram consists of three plates and an aperture stop. Two of these plates are virtual ; they are the missing anastigmatizing plate P_1 of the primary M_1, and the image P_2 in M_1 of the missing anastigmatizing plate of M_2 (both with strengths reversed, since the plates are missing). The third plate P_3 is the glass aspheric plate P of the system.

In the aplanats, the plate-strengths B_1, B_2, B_3 and the distances σ_1, σ_2, σ_3 of the plates in front of the aperture-stop satisfy the relations *

$$B_1 + B_2 + B_3 = 0, \qquad B_1\sigma_1 + B_2\sigma_2 + B_3\sigma_3 = 0, \qquad \ldots\ldots(11)$$

while the off-axis astigmatism is proportional to

$$B_1\sigma_1{}^2 + B_2\sigma_2{}^2 + B_3\sigma_3{}^2. \qquad \ldots\ldots(12)$$

Taking M_1 as defining the axis of the system by means of the normal through its pole, we set $\rho_1 = 0$ in the general formulae (8), (9), which then become, for an off-axis beam of azimuth χ,

Coma :

$$-S^2[2B_2\rho_2 e^{i\phi_2} + 2B_3\rho_3 e^{i\phi_3} + e^{2iE}(B_2\rho_2 e^{-i\phi_2} + B_3\rho_3 e^{-i\phi_3})]. \qquad \ldots\ldots(13)$$

Astigmatism and focus-shift :

$$S[2e^{iE}(B_2\rho_2{}^2 + B_3\rho_3{}^2) + e^{-iE}(B_2\rho_2{}^2 e^{2i\phi_2} + B_3\rho_3{}^2 e^{2i\phi_3})]$$
$$- 2SV[2e^{iE}\{B_2\rho_2\sigma_2 \cos(\phi_2 - \chi) + B_3\rho_3\sigma_3 \cos(\phi_3 - \chi)\}$$
$$+ e^{-iE+i\chi}\{B_2\rho_2\sigma_2 e^{i\phi_2} + B_3\rho_3\sigma_3 e^{i\phi_3}\}]$$
$$+ SV^2(2e^{iE} + e^{2i\chi - iE})(B_1\sigma_1{}^2 + B_2\sigma_2{}^2 + B_3\sigma_3{}^2); \qquad \ldots\ldots(14)$$

as already remarked, the terms Se^{iE} represent the focus-shift and those in Se^{-iE} the astigmatism.

In the lining-up procedure described above, the decentring (ρ_3, ϕ_3) given to the plate P_3 is determined by the condition that the coma shall vanish, i.e. by the equation

$$B_2\rho_2 e^{i\phi_2} + B_3\rho_3 e^{i\phi_3} = 0,$$

* The notation here and below is that of a previous paper (*Mon. Not. R. Astr. Soc.* **104**, 48 (1944)). Since the unit of plate-strength is there defined as the strength of the Schmidt plate of M_1, it follows that expressions (13) and (14) below must be multiplied by $f/8f_1{}^3$ to obtain their contributions to the actual displacements in the Petzval surface, and that the value of B_1 is -1.

which is equivalent, since B_2 and B_3 are both positive, to the relations

$$\phi_3 = \pi + \phi_2, \qquad B_3\rho_3 = B_2\rho_2. \qquad \dots \dots (15)$$

Expression (14) then reduces to

$$S(2e^{iE} + e^{2i\phi_2 - iE})\frac{B_2}{B_3}\rho_2{}^2$$

$$- 2SV(\sigma_2 - \sigma_3)[2e^{iE}\cos(\phi_2 - \chi) + e^{-iE + i\phi_2 + i\chi}]B_2\rho_2$$

$$+ SV^2(2e^{iE} + e^{2i\chi - iE})(B_1\rho_1{}^2 + B_2\sigma_2{}^2 + B_3\rho_3{}^2), \qquad \dots \dots (16)$$

in which $B_1 = -1, B_2 + B_3 = 1$.

If the plate and mirrors are free from errors of figure, the reduction of the observed astigmatism in the axial image to zero by tilting M_2 (operation (d)) is equivalent to setting $\rho_2 = 0$ in (16), i.e. to the correct alignment of M_2 relative to M_1. It then follows from (15) that $\rho_3 = 0$ and the system is correctly centred.

The practically important case, however, is that in which the secondary mirror M_2 is slightly out of alignment in relation to the primary; that is to say, the virtual plate P_2 is decentred and the consequent decentring of P_3 is given by (15). The decentring of P_2 may represent the result of applying operation (d) to a system whose plate is not quite free from astigmatism of figure, or it may represent an inaccuracy in the mechanical positioning of M_2 relative to M_1. In either case, we only need to consider lateral shifts of P_2, since the longitudinal adjustment of M_2 is made empirically.*

To fix ideas, we consider the special case of the two-sphere one-plate astrographic camera

$$q = 0.30, \qquad A = 0, \qquad B = 0, \qquad \Gamma = 0.7399, \qquad \sigma = 2.1447, \quad \dots \dots (17)$$

constructed with $f_1 = f_2 = 20$ inches. We suppose that the aperture stop, 10 inches in diameter, is situated 20 inches in front of the pole of the primary mirror† and that the system, of focal length approximately 28 inches, covers a field of 6° diameter at a nominal aperture of $F/3$. Then in the plate-diagram of the system

$$B_1 = -1, \qquad B_2 = q^2(2-q)^2 = 0.2601, \qquad B_3 = \Gamma = 0.7399;$$

$$\sigma_1 = 1, \qquad \sigma_2 = \frac{3-q}{2-q} - 1 = 0.5882, \qquad \sigma_3 = \sigma - 1 = 1.1447, \quad \dots \dots (18)$$

and the unit of length $f_1 = 20$ inches.

To express the effect of decentring P_2 in the azimuth $\phi_2 = 0$ and readjusting P_3 to annul coma all over the field, we set

$$\phi_2 = 0, \qquad \rho_3 = \frac{B_2}{B_3}\rho_2 = 0.3515\rho_2, \qquad \phi_3 = \pi.$$

Expression (16) then becomes

$$S(2e^{iE} + e^{-iE}) \times 0.3515\rho_2{}^2 + SV(2e^{iE}\cos\chi + e^{-iE + i\chi}) \times 0.2895\rho_2$$

$$+ SV^2(2e^{iE} + e^{2i\chi - iE}) \times 0.0596,$$

---

* An error in the longitudinal adjustment of M_2 would cause a change in the strength of P_2 as well as in its position.

† The stop is placed about half-way between the plate and the primary mirror in order to obtain uniform lighting of the field with the least possible increase in the diameters of plate and primary over that of the stop.

since

$$B_1\sigma_1{}^2 + B_2\sigma_2{}^2 + B_3\sigma_3{}^2 = \frac{q^2(1-q)^2}{B_3} = 0.0596.$$

The actual displacement-contribution in the Petzval surface for the ray through the point (S, E) of the aperture stop is therefore,

$$\delta x + i\delta y = \frac{f}{8f_1{}^3}(\Delta x + i\Delta y)$$

$$= \frac{S}{8(1-q)}[(2e^{iE} + e^{-iE}) \times 0.3515\,\rho_2{}^2/f_1{}^2$$

$$+ (2e^{iE}\cos\chi + e^{-iE+\chi}) \times 0.2895\,V\rho_2/f_1$$

$$+ (2e^{iE} + e^{2i\chi-iE}) \times 0.0596\,V^2], \qquad \dots\dots(20)$$

where V denotes the off-axis angle of the incident pencil and χ its azimuth.

In the case $\rho_2 = 0$ of a correctly lined-up system, the contribution of the off-axis astigmatism to the displacement is

$$(2e^{iE} + e^{2i\chi-iE}) \times 0.0596\,SV^2/8(1-q).$$

Since the absolute value of this expression never exceeds

$$3 \times 0.06 \times 5 \times \left(\frac{1}{20}\right)^2 \times \frac{10}{56} = 0.0004 \text{ inch,}$$

we see that the system is for all practical purposes anastigmatic.

In the case $\rho_2 > 0$ we have at the edge of the field:

(Change in $\delta x + i\delta y$ due to decentring)

$$= \frac{S}{8(1-q)}[(2e^{iE} + e^{-iE}) \times 0.00072\,\rho_2$$

$$+ (2e^{iE}\cos\chi + e^{-iE+i\chi}) \times 0.00072\,\rho_2],$$

where ρ_2, S are measured in inches. Since $S \leqslant 5$, the absolute value of this expression never exceeds

$$0.001(2.35\,\rho_2{}^2 + 1.94\,\rho_2) \text{ inch.}$$

A decentring in which ρ_2 does not exceed 0.1 therefore disturbs the contribution from astigmatism and focus-shift to the Seidel displacement $\delta x + i\delta y$ in the Petzval surface by an amount which nowhere exceeds 0.00022 inch in absolute value. Its effect on the image-quality will therefore be harmlessly small.

There is, however, another consequence of the decentring which needs to be considered, namely a disturbance in the distortion of the system, called " dislocation of the image" by Conrady. Only if this also is harmlessly small can the decentring be tolerated in a system whose images are to be used for measurements of position. To settle this point, we use expression (10), which in the case $\phi_2 = 0$, $\phi_3 = \pi$, $B_2\rho_2 = B_3\rho_3$ becomes

$$(B_2\rho_2{}^3 - B_3\rho_3{}^3) - V(B_2\rho_2{}^2\sigma_2 + B_3\rho_3{}^2\sigma_3)(e^{i\chi} + 2\cos\chi)$$

$$+ V^2(B_2\rho_2\sigma_2{}^2 - B_3\rho_3\sigma_3{}^2)(2e^{i\chi}\cos\chi + 1) - V^3 e^{i\chi}\sum_r B_r\sigma_r{}^3.$$

The first term merely represents a lateral displacement of the whole field, while the last term is the ordinary plate-contribution to the Seidel distortion of the

system.   The effect of the decentring is represented by the terms in V and V^2, whose contribution to the actual displacement $\delta x + i\delta y$ in the Petzval surface is

$$\frac{f}{8f_1^3}\left[-VB_2\rho_2^2\left(\sigma_2 + \frac{B_2}{B_3}\sigma_3\right)\left(e^{i\chi} + 2\cos\chi\right)\right.$$
$$\left. + V^2B_2\rho_2(\sigma_2^2 - \sigma_3^2)(2e^{i\chi}\cos\chi + 1)\right]$$
$$= \frac{1}{2240}[-0.2576\, V\rho_2^2(2e^{i\chi} + e^{-i\chi}) - 0.2508\, V^2\rho_2(2e^{2i\chi} + 1)].$$

Since $V \leqslant \dfrac{1}{20}$, the absolute value in inches of this contribution never exceeds

$$(17.5\rho_2^2 + \rho_2)10^{-5},$$

which for $\rho_2 \leqslant 0.1$ is utterly negligible.

It follows that a decentring for which $\rho_2 \leqslant 0.1$ causes no appreciable dislocation of the image, nor any serious deterioration in the image-quality.   Centring of this accuracy is readily obtainable by mechanical methods.*

Finally, we remark that the compensation of an axial astigmatism

$$\delta x + i\delta y = \tfrac{1}{2}Se^{-iE} \times 0.00022$$

due to error in the plate-figure can be shown by the same methods to involve a decentring, of approximately five times the above tolerance limit, which introduces astigmatism near the edge of the field several times greater than that removed at the centre.   The use of an intentionally decentred line-up is therefore not to be recommended.

REFERENCES .

BURCH, C. R., 1942.   *Mon. Not. R. Astr. Soc.* 102, 159.
CONRADY, A. E., 1919.   *Mon. Not. R. Astr. Soc.* 79, 384.
COX, H. W., 1939.   *J. Sci. Instrum.* 16, 257.
LINFOOT, E. H., 1944.   *Mon. Not. R. Astr. Soc.* 104, 48 ; 1945.   *Proc. Phys. Soc.* 57, 199.

* If it is more convenient to line-up the two mirrors optically, an adaptation of the method described by A. Couder (*Bull. Astronomique*, 1932, 7, 278–280) can be used with advantage.

THE FUNDAMENTAL FREQUENCY OF VIBRATION OF RECTANGULAR WOOD AND PLYWOOD PLATES

By R. F. S. HEARMON,

Princes Risborough

ABSTRACT. Equations are derived expressing the frequency of vibration of rectangular wood and plywood plates in terms of elastic constants, dimensions and density. For supported edges, the solution is exact and gives the complete series of overtones. For clamped edges, the fundamental frequency is estimated by the approximate Rayleigh method, but evidence is presented which indicates that the errors introduced by the use of the Rayleigh method are, in fact, negligible. A method of measuring the frequencies is described; comparison of calculated with observed frequencies shows that the former are on the average about 23 % high for clamped edges and about the same amount low for supported edges. This discrepancy is attributed to lack of correspondence between the experimental edge conditions and those assumed in the theoretical treatment. Finally, an empirical method, based on the rough proportionality between elastic constants and density of wood, is suggested and justified for obtaining approximate estimates of the fundamental frequencies in terms of plate dimensions only.

INTRODUCTION

CONSIDERABLE advances have recently been made in the elastic theory of wood and plywood. This theory has been used to calculate stresses and deflections under lateral loading and critical buckling stresses of plywood plates ; the agreement with experiment is reasonable though not exact. The frequency of vibration can be obtained theoretically by closely allied methods, and one of the objectives of the present investigation is to extend the comparison between theory and experiment to this frequency. A second objective is to gain some information regarding the frequency, which does not appear to have been studied hitherto.

§ 1. DEFINITIONS AND BASIC EQUATIONS

As is usual in problems of this type, it is assumed that the material is homogeneous and perfectly elastic and that, in the two-dimensional case, its elastic behaviour is determined by the existence of two perpendicular symmetry axes. For the particular purpose of the present investigation, it is further assumed that the symmetry axes (i.e. the directions parallel and perpendicular to the grain) are parallel with the sides of the plate. The dimensions of the plate are a, b and h, where

a = length of side parallel with Ox.

b =   „   „   „   „   „ Oy.

h = thickness of plate, parallel with Oz.

(h is small in comparison with a and b.)

Under these conditions the differential equation governing the small deflections w in the Oz direction is (March, 1936, 1942)

$$D_1 \frac{\partial^4 w}{\partial x^4} + D_2 \frac{\partial^4 w}{\partial y^4} + 2D_3 \frac{\partial^4 w}{\partial x^2 \partial y^2} - P = 0, \qquad \ldots \ldots (1.1)$$

where

$$\left.\begin{aligned}
D_1 &= \frac{E_1 h^3}{12\lambda} ; \\[2mm]
D_2 &= \frac{E_2 h^3}{12\lambda} ; \\[2mm]
D_3 &= \frac{E_1 \sigma_{21} h^3}{12\lambda} + \frac{G_{12} h^3}{6} \\[2mm]
&= D_4 + 2D_5 ;
\end{aligned}\right\} \qquad \ldots \ldots (1.2)$$

$E_1 = $ Young's modulus in bending in Ox direction ;

$E_2 = $ „   „   „   „ · „ Oy „   ;

$G_{12} = $ Rigidity modulus for shear stresses in Ox and Oy directions ;

$\sigma_{21} = \dfrac{\text{contraction in } Ox \text{ direction}}{\text{extension in } Oy \text{ direction}}$ for tension in Oy direction ;

$\lambda = 1 - \sigma_{21}\sigma_{12} \approx 0.99$ for wood and plywood ;

$P = $ pressure acting in Oz direction.

The differential equation governing the vibration of the plate can be found by substituting $\rho h (\partial^2 w / \partial t^2)$ for $-P$ in (1.1) :

$$D_1 \frac{\partial^4 w}{\partial x^4} + D_2 \frac{\partial^4 w}{\partial y^4} + 2D_3 \frac{\partial^4 w}{\partial x^2 \partial y^2} + \rho h \frac{\partial^2 w}{\partial t^2} = 0, \qquad \ldots \ldots (1.3)$$

where $\rho = $ density of material (weight at test/volume at test) and $t = $ time.

The boundary conditions for a plate clamped at the edges are (Timoshenko, 1940, § 22) :

$$\left.\begin{aligned}
w &= 0 \text{ at :---} & x &= 0, & x &= a, & y &= 0, & y &= b ; \\[2mm]
\frac{\partial w}{\partial x} &= 0 \text{ at :---} & x &= 0, & x &= a ; & & & & \\[2mm]
\frac{\partial w}{\partial y} &= 0 \text{ at :---} & y &= 0, & y &= b; & & & &
\end{aligned}\right\} \qquad \ldots \ldots (1.4)$$

while for a plate simply supported at the edges they are (Timoshenko, 1940, §§ 22, 26) :

$$\left.\begin{aligned}
w &= 0 & \text{at :---} & & x &= 0, & x &= a, & y &= 0, & y &= b ; \\[2mm]
\frac{\partial^2 w}{\partial x^2} &= 0 & \text{at :---} & & x &= 0, & x &= a ; & & & & \\[2mm]
\frac{\partial^2 w}{\partial y^2} &= 0 & \text{at :---} & & y &= 0, & y &= b. & & & &
\end{aligned}\right\} \qquad \ldots \ldots (1.5)$$

The potential energy of bending (V) is (March, 1942 ; Goland, 1942) :

$$V = \frac{1}{2}\int_0^a\int_0^b \left\{ D_1\left(\frac{\partial^2 w}{\partial x^2}\right)^2 + D_2\left(\frac{\partial^2 w}{\partial y^2}\right)^2 + 2D_4\frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2} + 4D_5\left(\frac{\partial^2 w}{\partial x\partial y}\right)^2 \right\}dx\,dy,$$

$$\ldots\ldots(1\cdot6)$$

and the kinetic energy T is

$$T = \frac{\rho h}{2}\int_0^a\int_0^b\left(\frac{\partial w}{\partial t}\right)^2 dx\,dy. \qquad \ldots\ldots(1\cdot7)$$

§2. DERIVATION OF THE FREQUENCY EQUATIONS

In the case of supported edges, an exact solution can be obtained by expressing w as

$$w = w_0\sin\frac{m\pi x}{a}\sin\frac{n\pi y}{b}\sin pt, \qquad \ldots\ldots(2.1)$$

where $p = 2\pi \times$ frequency (v), m and n are integers, and w_0 is a constant. Substitution of (2.1) in (1.5) shows that the boundary conditions are satisfied, and substitution of (2.1) in the differential equation·(1.3) leads directly to

$$v_s = \frac{\pi}{2a^2b^2}\left\{\frac{1}{\rho h}[D_1 m^4 b^4 + D_2 n^4 a^4 + 2D_3 m^2 n^2 a^2 b^2]\right\}^{\frac{1}{2}}, \qquad \ldots\ldots(2.2)$$

where the suffix s indicates supported edges.

The fundamental is given by $m = n = 1$, and the overtones by higher values of m and n.

In view of the complexity of the problem involved in a complete solution, the fundamental frequency for clamped edges has been derived by the Rayleigh method (see e.g. Timoshenko, 1937, § 16), supplemented to some extent by the use of the more accurate Ritz modification. For comparison both with the exact solution (2.2) and with the results for clamped edges, the Rayleigh and Ritz methods have also been applied to the case of simply supported edges.

The Rayleigh method consists in working out the potential energy V and the kinetic energy T from the equation

$$w = f(x, y)g(t), \qquad \ldots\ldots(2.3)$$

which expresses the deflection w as a function of time t and of position on the plate. If the system is conservative, then

$$V_{\text{max}} = T_{\text{max}}, \qquad \ldots\ldots(2.4)$$

and the frequency of vibration is found from this condition.

The virtue of the Rayleigh method is that in general it is possible to use an approximate expression of $f(x, y)$ in (2.3) and still obtain estimates of the frequency sufficiently accurate for all practical purposes. In the present instance, the simplest algebraic expressions consistent with the boundary conditions have been chosen to represent $f(x, y)$. Assuming a sinusoidal vibration, this leads to

$$w = K_c x^2 y^2 (a - x)^2 (b - y)^2 \sin pt, \qquad \ldots\ldots(2.5)$$

for clamped edges, and to

$$w = K_s xy(a - x)(b - y)(a^2 + ax - x^2)(b^2 + by - y^2)\sin pt, \qquad \ldots\ldots(2.6)$$

for supported edges, where K_c and K_s are constants, the suffix indicating the boundary conditions. It should be noted that, although (2.5) and (2.6) satisfy the boundary conditions, they are not solutions of (1.3).

The potential energy can now be found for clamped edges by inserting w, as given by (2·5), in (1·6) :

$$V_c = \frac{a^5 b^5 K_c{}^2}{1575}\left[D_1 b^4 + D_2 a^4 + \frac{4}{7}D_3 a^2 b^2\right]\sin^2 pt, \qquad \ldots\ldots(2.7)$$

and for supporting edges by inserting (2.6) in (1.6) :

$$V_s = \frac{62 a^5 b^5 K_s{}^2}{525}\left[D_1 b^4 + D_2 a^4 + \frac{867}{434}D_3 a^2 b^2\right]\sin^2 pt. \qquad \ldots\ldots(2.8)$$

The kinetic energy is given similarly for clamped edges by (2.5) and (1.7) as

$$T_c = \frac{\rho h a^9 b^9 K_c{}^2 p^2 \cos^2 pt}{793800}, \qquad \ldots\ldots(2.9)$$

and for supported edges by (2.6) and (1.7) as

$$T_s = \frac{961 \rho h a^9 b^9 K_s{}^2 p^2 \cos^2 pt}{793800}. \qquad \ldots\ldots(2.10)$$

In each case the potential energy can be written $V = MK^2 \sin^2 pt$ and the kinetic energy $T = NK^2 p^2 \cos^2 pt$. The maximum V occurs when $\sin^2 pt = 1$ and the maximum T when $\cos^2 pt = 1$. Hence from condition (2.4), $p^2 (= 4\pi^2 \nu^2) = M/N$. Thus from equations (2.7) and (2.9)

$$\nu_c = \frac{1}{\pi a^2 b^2}\left\{\frac{126}{\rho h}\left[D_1 b^4 + D_2 a^4 + \frac{4}{7}D_3 a^2 b^2\right]\right\}^{\frac{1}{2}}, \qquad \ldots\ldots(2.11)$$

and from (2.8) and (2.10)

$$\nu_s = \frac{1}{\pi a^2 b^2}\left\{\frac{756}{31\rho h}\left[D_1 b^4 + D_2 a^4 + \frac{867}{434}D_3 a^2 b^2\right]\right\}^{\frac{1}{2}}. \qquad \ldots\ldots(2.12)$$

Comparison of equation (2.12) with the accurate solution (2.2), putting $m = n = 1$, shows that the form of the equations is identical, and further, the numerical coefficients are very close in the two cases. The coefficient outside the brackets in the accurate equation (2.2) is $\pi/2 = 1\cdot5708$, while the corresponding coefficient in (2.12) is $\pi^{-1}\sqrt{(756/31)} = 1\cdot5719$; the coefficient of the term in $D_3 a^2 b^2$ is 2 in equation (2.2) and $867/434 (= 1\cdot998)$ in equation (2.12).

In order to obtain an estimate of the degree of approximation involved in the clamped-edge solution (2.11), the Ritz modification of the Rayleigh method has been applied to the present problem. The procedure, with an example on the one-dimensional case of the vibrating string, is given in Timoshenko's book (1937, § 62). The application of the method to the two-dimensional case of the vibrating anisotropic plate does not present any special difficulties, although it leads to somewhat laborious algebra and arithmetic. For the above reasons, only the final results, and such description of the process as is required for their presentation, will be given here.

6

The forms assumed for w are generalizations of (2.5) and (2.6) :

(1) Clamped edges, $w = X^2 Y^2 (K_1 + K_2 XY) \sin pt$;

(2) Supported edges, $w = XY(a^2 + ax - x^2)(b^2 + by - y^2)(K_1 + K_2 XY) \sin pt$,
where $X = x(a - x)$, $Y = y(b - y)$.

If $K_2 = 0$, the above equations reduce respectively to (2.5) and (2.6). An essential feature of the Ritz method is that the values of K_1 and K_2 are adjusted at one stage in the process so as to make the frequency a minimum. With two constants, this leads to a quadratic in the frequency, one root of which gives a more accurate estimate of the fundamental frequency than that given by the Rayleigh method.

For the purpose of comparing the various estimates, it is convenient to confine attention to square plates, and, in the case of wood and plywood, to particular values of the elastic constants. For square plates, equations (2.2), (2.11) and (2.12) are of the form

$$v = c_1 / a^2 \sqrt{\rho h}. \qquad \qquad \cdots\cdots (2.13)$$

Values of the constant c_1 have been calculated by the Ritz method for plates of birch 5-ply and birch veneer having the following " flexural rigidities " D (c.g.s. units ; values based on elastic constants given in table 2 below).

	D_1	D_2	D_3
Birch 5-ply	$0 \cdot 191 \times 10^{10}$	$0 \cdot 071 \times 10^{10}$	$0 \cdot 044 \times 10^{10}$
Birch veneer	$2 \cdot 97 \times 10^8$	$0 \cdot 21 \times 10^8$	$0 \cdot 69 \times 10^8$

For an isotropic solid $D_1 = D_2 = D_3$, and it is therefore possible, by introducing this condition, to reduce any of the equations so far given to the isotropic case. The exact value of c_1 for a square isotropic plate with supported edges is known (Timoshenko, 1937, equation 214) to be $\pi \sqrt{D_1}$, where $D_1 = Eh^3 / 12(1 - \sigma^2)$, and

Table 1. Calculated values of c_1 in equation (2.13)
(Square plates)

Method	Clamped edges			Supported edges		
	Isotropic	Birch		Isotropic	Birch	
		5-ply	Veneer		5-ply	Veneer
Rayleigh	$5 \cdot 730 \sqrt{D_1}$	$1 \cdot 914 \times 10^5$	$0 \cdot 6755 \times 10^5$	$3 \cdot 1429 \sqrt{D_1}$	$0 \cdot 9300 \times 10^5$	$0 \cdot 3357 \times 10^5$
Ritz	$5 \cdot 729 \sqrt{D_1}$	$1 \cdot 912 \times 10^5$	$0 \cdot 6750 \times 10^5$	$3 \cdot 1420 \sqrt{D_1}$	$0 \cdot 9294 \times 10^5$	$0 \cdot 3355 \times 10^5$
Exact	$5 \cdot 727 \sqrt{D_1}$	—	—	$3 \cdot 1416 \sqrt{D_1}$	$0 \cdot 9293 \times 10^5$	$0 \cdot 3354 \times 10^5$

Tomotika (1936) has given an estimate of c_1 for a square plate with clamped edges which, though not strictly exact, gives a highly accurate lower limit. The various estimates for c_1 are collected in table 1.

Even the largest difference in the above table is only just greater than $0 \cdot 1\%$, and it can be concluded that the fundamental frequencies estimated by the

Rayleigh method are mathematically sufficiently accurate for the purpose of comparing theoretical with observed frequencies. The Rayleigh equations (2.11) and (2.12) have, therefore, been used in all subsequent calculations.

§3. EXPERIMENTAL

The general features of the method used to determine the frequencies can be seen from figures 1, 2 and 3. The wood or plywood plate was held in a suitable frame and two small iron plates A and C (figure 1) were attached by means of soft wax, one (A) at the centre of the test plate, and the other (C) a convenient distance away. A small iron-cored solenoid (B), fed by alternating current from a beat-frequency oscillator, was placed near to plate A, while a telephone receiver, stripped of its diaphragm and case, was placed so that the magnets were close to plate C. Any currents induced in the coils of the telephone were fed to a cathode-ray oscillograph, and the experiment consisted in varying the frequency of the driving current supplied to B until the response, as indicated by the oscillograph, was a maximum. The telephone receiver D was connected to the vertical plates of the oscillograph, and the driving current from the oscillator connected to the horizontal plates, in parallel with the coil B. In this way, a

FROM OSCILLATOR
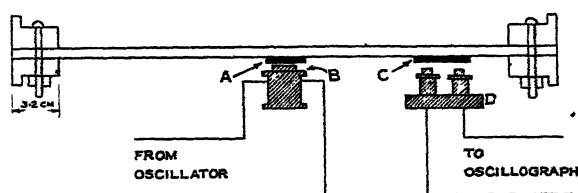
TO OSCILLOGRAPH

Figure 1. Experimental arrangement.

Lissajous figure was obtained on the oscillograph screen, and it was found that, with the present experimental arrangement, the plates tended to respond to integral submultiples of the resonant frequency, in agreement with results, so far unpublished, on wooden strips. The most powerful response in all cases occurred when the frequency of the driving current was one-half the true resonant frequency, as is shown by the fact that the Lissajous figure indicated two vibrations of the plate for each cycle of the forcing current. The results given in table 2 are all based on this response mode.

The apparatus can, of course, be used to measure overtones, and a number of observations were actually made of the frequencies at which these occurred. Owing to certain difficulties, mainly theoretical, a satisfactory interpretation of these measurements is not available at present, and they will not be discussed further.

The frames for holding the plates were made of wood and are shown in figures 1, 2 and 3. The frame for clamped edges (general view, figure 2; cross-section, figure 1) consisted of two halves bolted together, with the plate sandwiched between them. The frame for supported edges (figure 3) consisted of four separate grooved members held together by brass bolts and wing nuts, with the edges of the plate at the base of the grooves, as shown in the section diagram, figure 3 (*b*).

For convenience of lining up and adjustment, the apparatus was assembled on an optical-bench base.   The frames were held at opposite corners by ordinary laboratory clamps attached to vertical rods ; the solenoid B and telephone receiver D were also mounted on rods, all of which were held in optical-bench stands.

The plates were originally cut to 45·8 cm. squares, and were progressively reduced in size in order to provide measurements over a range of lateral dimensions.   It can be seen from equations (2.11) and (2.12) that if D_1 and D_2 differ,
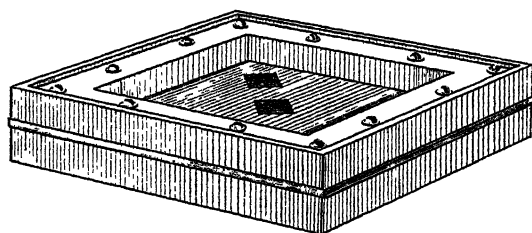
Figure 2.   Clamped edge frame.

the frequency of an oblong plate of given dimensions depends on whether the direction of greater flexural rigidity coincides with the longer or shorter side of the plate, and provision was made for the testing of plates fulfilling both of these conditions.   For example, two 45·8 × 45·8 cm. plates of obeche veneer were originally available (table 2).   The first reduction in size was accomplished on plate 1 by shortening to 30·6 cm. the side *a*, lying parallel to the greater Young's modulus, and on plate 2 by shortening to the same length the side *b*, lying parallel
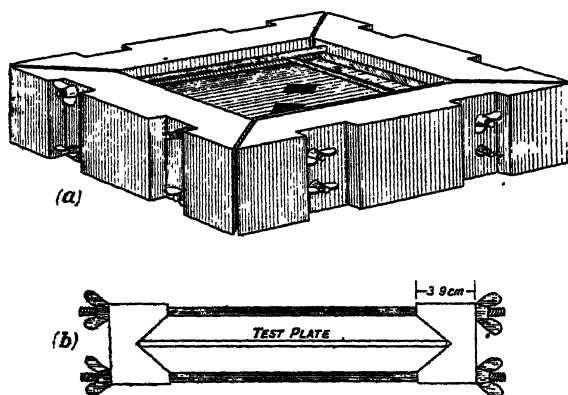
Figure 3.   Supported edge frame.
(a) General view ;   (b) cross-section.

to the smaller Young's modulus.   On the second reduction in size, side *b* on plate 1 and side *a* on plate 2 were shortened to 30·6 cm., thus giving square plates again.   This procedure was repeated once more, finally leading to the smallest size of plate, 20·4 × 20·4 cm.   The dimensions given refer to the supported-edge condition ; the clamped dimensions are smaller owing to the overlap of the frames (see figures 1 and 2).

Young's moduli in bending (*E*), and rigidity modulus (*G*), on which, by equations (1.2), the " flexural rigidities " depend, were determined respectively

| | Veneers | | | | | | | | Plywoods | | | | |
	Obeche (1)	Obeche (2)	Gaboon (1)	Gaboon (2)	Gaboon (3)	Beech (1)	Beech (2)	Birch	Birch 5-ply	Birch 3-ply (1)	Birch 3-ply (2)	Opepe 3-ply	Douglas fir 3-ply
(1) Elastic constants													
E_1 $\times 10^{-10}$ $\dfrac{\text{dynes}}{\text{sq. cm.}}$	6.9	7.4	7.8	7.9	8.7	13.3	13.3	11.7	9.3	12.0	12.4	9.3	8.4
E_2	0.17	0.05	0.30	0.19	0.14	0.55	0.54	0.82	3.45	1.19	1.21	0.80	1.08
$E_1\sigma_{21}$	0.17	0.17	0.19	0.19	0.22	0.33	0.33	0.37	0.37	0.37	0.37	0.25	0.25
G_{12}	0.30	0.34	0.58	0.45	0.41	0.85	0.86	1.19	0.90	0.88	0.93	0.54	0.90
h (cm.)	0.291	0.323	0.318	0.310	0.315	0.309	0.309	0.311	0.626	0.532	0.525	0.479	0.415
ρ (gm./cc.)	0.330	0.390	0.400	0.399	0.393	0.670	0.659	0.719	0.690	0.716	0.711	0.774	0.533
(2) Observed fundamental frequencies (cycles/sec.). Clamped edges													
$a=b=39.3$ cm.	67	82	76	76	78	89	82	—	160	140	148	—	—
$a=24.2,\ b=39.3$	172	—	192	—	—	95	182	96	—	—	—	—	—
$a=39.3,\ b=24.2$	—	88	—	96	92	—	—	—	254	197	254	155	162
$a=b=24.2$ cm.	189	191	202	156	202	200	225	212	392	320	326	270	278
$a=14.0,\ b=24.2$	506	—	496	—	—	—	440	—	650	486	706	375	382
$a=24.2,\ b=14.0$	—	201	—	229	224	241	—	248	—	—	—	—	—
$a=b=14.0$ cm.	444	502	556	510	524	530	514	544	980	830	850	654	618
(3) Observed fundamental frequencies (cycles/sec.). Supported edges													
$a=b=45.8$ cm.	35	39	40	34	38	41	43	—	73	67	61	—	—
$a=30.6,\ b=45.8$	92	—	94	—	—	—	95	52	—	—	—	—	—
$a=45.8,\ b=30.6$	—	60	—	68	50	53	—	—	150	82	152	77	90
$a=b=30.6$ cm.	81	114	105	88	92	84	94	108	187	190	180	120	137
$a=20.4,\ b=30.6$	186	—	181	—	—	—	166	—	281	200	320	167	168
$a=30.6,\ b=20.4$	—	94	—	109	92	106	—	123	—	—	—	—	—
$a=b=20.4$ cm.	184	190	218	182	182	186	191	228	420	390	374	296	278

by flexural and torsional vibration experiments (Hearmon and Barkas, 1941)
on approximately 2·5-cm. wide strips of the material cut parallel to and near
the sides of the largest plates. The Poisson's ratios (σ) were not measured, but
were obtained from published values of σ for the appropriate solid wood. In
some cases, no relevant measurements could be found, and it was then assumed
that the Poisson's ratio for solid wood, $\dfrac{\text{contraction along the grain}}{\text{extension across the grain}}$ for tension
across the grain, had the value 0·025. As in much work on wood and plywood,
the contribution of the terms involving the Poisson's ratios is small, and the
exact values adopted are not very important.

§4. RESULTS AND DISCUSSION

The complete experimental results are given in table 2 ; the side a is taken
as parallel with the direction of E_1 and b with E_2, where $E_1 > E_2$.

In order to compare the observed and theoretical frequencies, it is convenient
to express the equations for the latter in terms of elastic constants, rather than
flexural rigidities, by substituting equations (1.2) in (2.11) and (2.12). Thus,

$$\nu_c = \frac{1\cdot04h}{a^2\sqrt{\rho}}\left[E_1 + \left(\frac{a}{b}\right)^4 E_2 + \left(\frac{a}{b}\right)^2 (0\cdot57 E_1\sigma_{21} + 1\cdot13 G_{12})\right]^{\frac{1}{2}} \quad \ldots\ldots(4.1)$$

and

$$\nu_s = \frac{0\cdot456h}{a^2\sqrt{\rho}}\left[E_1 + \left(\frac{a}{b}\right)^4 E_2 + \left(\frac{a}{b}\right)^2 (2\cdot00 E_1\sigma_{21} + 3\cdot96 G_{12})\right]^{\frac{1}{2}}, \quad \ldots\ldots(4.2)$$

where λ has been put equal to 0·99, and the numerical coefficients are given to
three significant figures.

The observed frequencies are plotted against those calculated from (4.1) and
(4.2) in figures 4 and 5. The lines in these figures are not necessarily the "best"
fit, but are drawn through the origin with slope determined by the average
value of the ratio (observed frequency/calculated frequency) for all the obser-
vations. Numerically, this average is 0·772 for clamped edges and 1·238 for
supported edges. Thus, the calculated frequencies with edges clamped are
consistently higher than the observed frequencies, while the reverse is true for
the supported edges. This suggests that the actual conditions at the edges of
the plates are always intermediate between the conditions appropriate to strictly
clamped and supported edges (equations (1.4) and (1.5)). The difficulty of
realizing experimentally the exact boundary conditions assumed in the theoretical
treatment of this type of problem is well known, and has been encountered, for
example, in buckling tests of plywood panels in shear (Norris, Voss and Palma,
1945).

Ideally, of course, the points in figures 4 and 5 should lie on a line passing
through the origin, and having a slope of 45° ; the general tendency of the points
to depart from this line is a measure of the effect just discussed. In addition,
there is also a certain amount of random scatter, which is partly accounted for by
casual experimental errors in the frequency measurements, or by variation in
the material between the test plates and the strips used in the determination of
elastic constants.

Superimposed on all these effects, however, there is definite indication that with clamped edges (figure 4) the best line through the points is curved, and not straight ; the curvature is such that the ratio (ν observed/ν calculated) becomes smaller at higher frequencies. The physical interpretation of this result is that the frame used for clamped edges becomes less effective with thicker plates and smaller lateral dimensions, and under these conditions departures from the ideal clamped-edge conditions tend to be higher.

In order to obtain further information on this effect, a mild steel frame was made to hold the smallest square plates in the clamped-edge condition. This frame was of the same dimensions as the corresponding wooden frame ; the two halves, however, were held together by twelve ¼-in. bolts instead of the ⅜-in. size used with the wooden frame. The observed frequencies in the steel and wooden frames and the theoretical frequencies (equation (4.1)) are given in table 3.
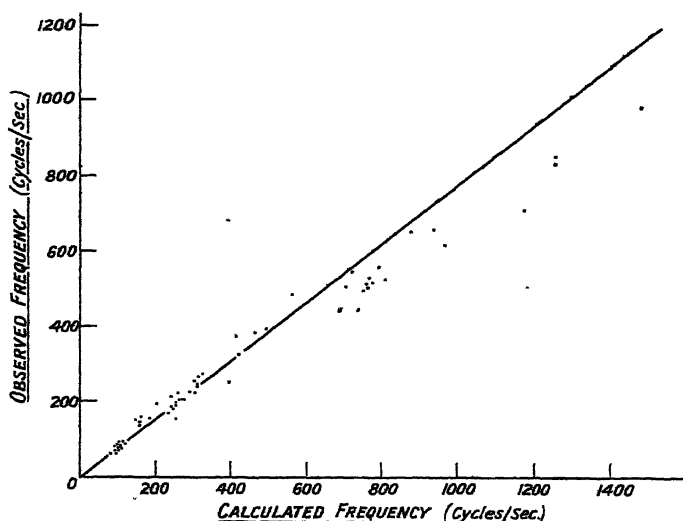
Figure 4. All results, clamped edges.

In each case the frequency is higher in the steel frame than in the wooden frame, and is, therefore, closer to the calculated frequency. The average ratio (observed frequency/calculated frequency) is 0·67 for the wooden frame and 0·83 for the mild-steel frame. These results confirm that a considerable part at least of the discrepancy between observed and calculated frequencies can be attributed to the conditions at the edges of the plates.

With the exception of the opepe 3-ply, the observed frequencies in the steel frame are lower than those calculated. This suggests that, even in the steel frame, the restraint at the edges is not sufficient to realize fully the ideal clamped-edge conditions.

Inspection of figure 5 (simply supported edges) does not reveal any marked tendency to curvature, though statistical analysis (see table 4) shows that there is in fact a slight tendency to curvature such that the ratio (observed frequency/calculated frequency) *increases* somewhat as the frequency rises. It is rather surprising that the supported-edge frame should prove more

R. F. S. Hearmon

Table 3. Frequencies in cycles/sec. Edges clamped ; $a = b = 14 \cdot 0$ cm.

Material	Obeche		Gaboon			Beech		Birch	Birch 5-ply	Birch 3-ply		Opepe 3-ply	Douglas fir 3-ply
	(1)	(2)	(1)	(2)	(3)	(1)	(2)			(1)	(2)		
Frequency (wooden frame)	444	502	556	510	524	530	514	544	980	830	850	654	618
Frequency (steel frame)	514	585	642	593	613	637	628	647	1220	1110	1110	960	770
Frequency (calculated)	740	770	800	770	810	780	790	730	1490	1260	1270	950	980

satisfactory in this respect than the clamped-edge frame, because the former has quite apparent defects. There is a small area near each corner which is unrestrained, but, perhaps more important, the degree to which the wing nuts holding the frame together (figure 3) are tightened is a matter for personal judgment in obtaining a sufficiently firm fixing of the test plate without applying appreciable compressions in its plane. Nevertheless, the experimental results indicate that, although the conditions at the edges do not constitute true simple support, they are to a first approximation independent of the dimensions of the plate.

In view of the rough proportionality between elastic constants and density of timber, it is possible to derive certain empirical relations which enable the frequencies to be estimated even if the elastic constants and densities are not
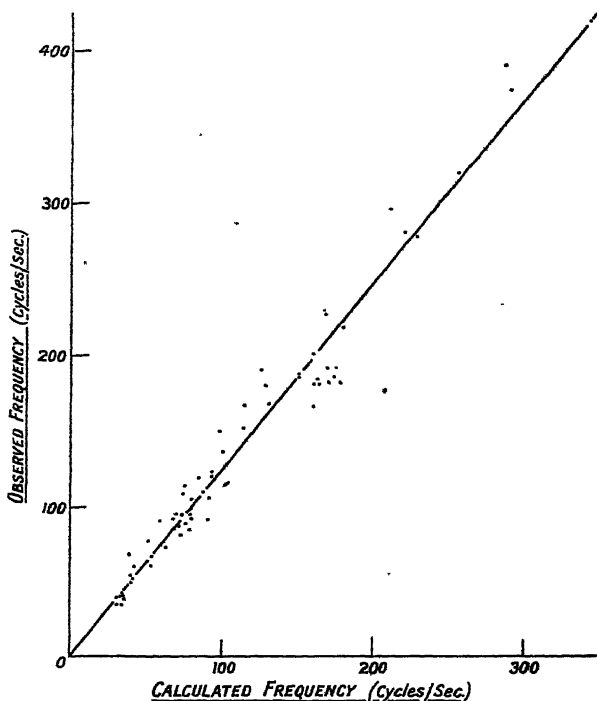
Figure 5. All results, supported edges.

known. If it is assumed that elastic constants are exactly proportional to density, then equations (4.1) and (4.2) take the form

$$\nu = c_2 h / a^2, \qquad\qquad \ldots\ldots (4.3)$$

where c_2 is a numerical factor which depends only on the edge conditions and on the ratio a/b.

Two related consequences follow from (4.3) :

(1) For given edge conditions, and given ratios a/b and h/a^2, the frequency should be constant irrespective of the values of elastic constants and density. Examination of the experimental results for the veneers in table 2, all of which have a thickness of about 0·3 cm., shows that this prediction is approximately verified.

(2) For given edge conditions, and a given value of a/b, a plot of ν against h/a^2 should give a straight line passing through the origin. The theoretical frequencies for square plates are plotted on this basis in figures 6 and 7. and it
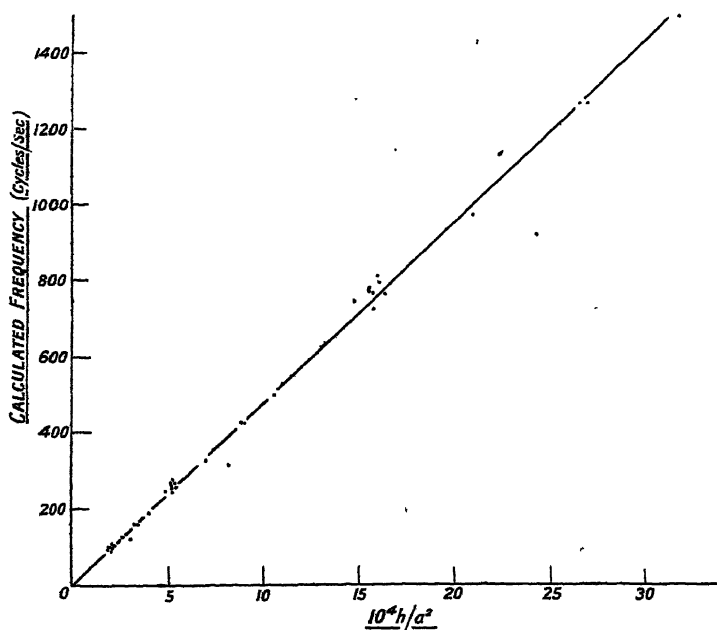
Figure 6.   Square plates, clamped edges.
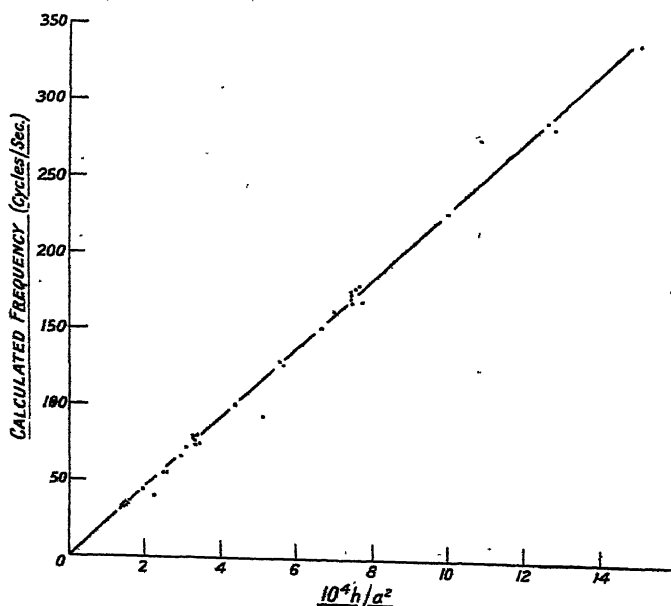
Figure 7.   Square plates, supported edges.

will be seen that, apart from three exceptionally low frequencies in each figure, the points lie very close to the lines

$$\nu_c = 47 \times 10^4 h/a^2 \text{ (clamped edges),} \qquad \ldots\ldots(4.4)$$
$$\nu_s = 23 \times 10^4 h/a^2 \text{ (supported edges).} \qquad \ldots\ldots(4.5)$$

The exceptionally low frequencies all refer to the opepe 3-ply, and serve as a reminder that this method of estimating frequencies must necessarily give results somewhat in error in those cases where the assumed proportionality between elastic constants and density does not hold.

The theoretical frequencies for oblong plates have been analysed in the same way and have given the following equations :

Clamped edges :
$$a/b = 2/3, \qquad v_c \approx 45 \times 10^4 h/a^2 ;$$
$$a/b = 3/2, \qquad v_c \approx 55 \times 10^4 h/a^2.$$

Supported edges :
$$a/b = 2/3, \qquad v_s \approx 21 \times 10^4 h/a^2 ;$$
$$a/b = 3/2, \qquad v_s \approx 29 \times 10^4 h/a^2.$$

It will be remembered that a is the length of side parallel to the direction of greatest Young's modulus. The scatter of the points for $a/b = 2/3$ is greater than that shown for square plates in figures 6 and 7, and the uncertainty in the predicted frequency is therefore greater in this case.

The experimental results for square plates (table 2) have been analysed statistically on the above basis, both as a whole and in groups corresponding with the different sizes of plate. The regression equations for predicting the experimentally observed frequencies in terms of $y(= 10^4 h/a^2)$ are summarized in table 4.

Table 4. Summary of regression equations
(v = observed frequency, $y = 10^4 h/a^2$)

Clamped edges		Supported edges	
$a = b = 39\cdot3$ cm.	$v = 42\cdot9y - 7\cdot3$	$a = b = 45\cdot8$ cm.	$v = 23\cdot6y + 3\cdot8$
$a = b = 24\cdot2$ cm.	$v = 34\cdot4y + 14\cdot9$	$a = b = 30\cdot6$ cm.	$v = 30\cdot7y - 6\cdot1$
$a = b = 14\cdot0$ cm.	$v = 27\cdot7y + 70\cdot2$	$a = b = 20\cdot4$ cm.	$v = 31\cdot8y - 44\cdot0$
All results	$v = 29\cdot3y + 38\cdot7$	All results	$v = 28\cdot3y - 2\cdot9$

For the largest sizes, the equations are not very different from (4.4) and (4.5), which are derived from the theoretical frequencies. With clamped edges there is a noticeable decrease in the slope of the lines and an increase in the intercept on the y axis with decreasing size of plate. These effects are reflections of the tendency to curvature already discussed in connection with figure 4. The supported-edge results show a slight tendency to curvature in the opposite direction when analysed in this way, although, as already mentioned, this tendency is not obvious on inspection of figure 5. In other words, the edge conditions in the frames tend to approach each other at the higher frequencies, and there are indications that at the highest frequencies the " clamped "-edge frame may even impose less restraint at the edges than the " supported "-edge frame.

ACKNOWLEDGMENTS

Research Board, and publication is by permission of the Departm ent of Scientific and Industrial Research. The statistical analyses summarized in table 4 .were carried out under the direction of Miss C. B. Pettifor, B.Sc.

REFERENCES

GOLAND, M., 1942. Curtiss-Wright Corporation, Airplane Division, *Research Report*, 5 S, 13.
HEARMON, R. F. S. and BARKAS, W. W., 1941. *Proc. Phys. Soc.* 53, 674.
MARCH, H. W., 1936. *Physics*, 7, 32.
MARCH, H. W., 1942. *U.S. Dept. Agric. Forest Prod. Lab., Madison*, Mimeo. 1312.
NORRIS, C. B., VOSS, A. W. and PALMA, J., 1945. *U.S. Dept. Agric. Forest Prod. Lab., Madison*, Mimeo. 1316 H.
TIMOSHENKO, S., 1937. *Vibration Problems in Engineering* (New York : Van Nostrand).
TIMOSHENKO, S., 1940. *The Theory of Plates and Shells* (New York : McGraw-Hill).
TOMOTIKA, S., 1936. *Phil. Mag.* 21, 745.

THE OPTICAL SINE-CONDITION

By H. H. HOPKINS,

W. Watson and Sons, Barnet, Herts

ABSTRACT. The Sine-Condition is formulated on the basis of optical path differences, leading to an expression which is valid in the presence of spherical aberration. It is shown to contain results obtained previously by means of a " ray-intersection " analysis. The " next terms " are estimated and serve as an approximate measure of the error involved ; or, alternatively, as an indication of the size of field over which the sine-condition is a valid measure of off-axis aberration.

IT is well known that fulfilment of the so-called *Sine-Condition* by an optical system ensures freedom from the aberration of coma for objects that are but a small angular distance from the axis. Because of this fact, the sine-condition has been used extensively by lens designers as a check on the off-axis performance of both telescope and microscope objectives. There are, however, respects in which this usage has been unsatisfactory.

In the first place the condition is a valid measure of off-axis aberration only when the angular distance off-axis of the object point is so small that its square can be neglected. There is an uncertainty in the size of field over which this approximation is permissible, given a required order. of accuracy in the result. Secondly, the physical meaning which can be attached to the violation of the condition is—in my experience—generally not appreciated. The permitted departures from exact fulfilment of the sine-condition are consequently either wholly arbitrary or based on experience of " what works ". It is true that Conrady (*Applied Optics and Optical Design*, vol. i, p. 393) gave a " Rayleigh tolerance " for coma, but this was simply quoted without proof.

. In what follows, an account of the imagery of off-axis object points is given which leads to a simple interpretation of departures from the sine-condition

and also provides an approximate rule giving the size of field over which the condition is a valid measure of off-axis aberration.

In figure 1, O_1 and Q_1 are object points (real or virtual), the one situated on, and the other a small angular distance off, the axis, O_1O_λ', of an optical system having λ refracting surfaces. The diagram shows a meridian section containing the off-axis object point Q_1. ST is a section of a spherical wave-front with centre O_1 : S_QT_Q is a section of a spherical wave-front with centre Q_1. Thus O_1P_0, $Q_1R_0P_0$ and O_1E_1, Q_1E_1 are rays along which elements of the waves proceed from the object points O_1, Q_1. R_0 is on the sphere S_QT_Q.

O_λ', Q_λ' are the Gaussian images of O_1, Q_1 respectively; $S'T'$, $S_Q'T_Q'$ are spheres centred on O_λ', Q_λ'. The forms of the wave-fronts emerging from the system and proceeding towards O_λ', Q_λ' can be specified by their departures at a given moment from the appropriate one of the spheres centred on O_λ', Q_λ'. Such a procedure has been used before.*
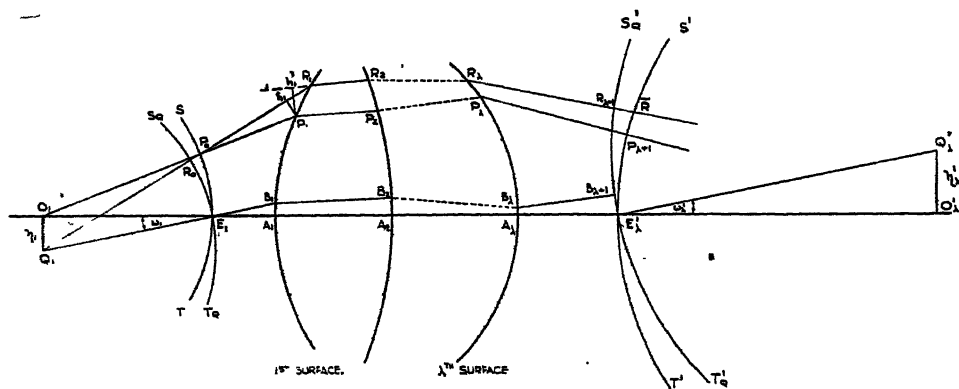
Figure 1.

When the departure from the appropriate reference sphere is zero (or constant) for the whole aperture, the corresponding image point is free from aberration. The departure at a given point on the wave-front will be referred to as the wave-front aberration. It is a function of the position of the wave element on the emergent wave-front, and will be measured, not as a geometrical, but as an optical, path-length.

The incident wave-fronts and the reference-spheres of the emergent wave-fronts are conveniently chosen so that they intersect in E_1, E_λ' the axial points of the entrance and exit pupils. In the diagram (simply for the purposes of clarity) these latter have been taken to be external to the system. Thus Q_1E_1 is the initially incident principal ray from Q_1.

An element of the wave from O_1 proceeds along the ray $O_1P_0P_1P_2\ldots\ldots P_\lambda P_{\lambda+1}$, intersecting the O_λ'-reference sphere in $P_{\lambda+1}$. Suppose the final wave-front aberration of this element is W_λ', measured positively when the element is in advance of the O_λ'-reference sphere, and such that this latter and the emergent wave coincide at E_λ'—that is, W_λ' is zero for the element of wave which proceeds

* For a derivation of the theory of the aberrations of lens-systems on this basis, see the writer's " Monochromatic Lens Aberration Theory ", *Phil. Mag.* (publication impending.)

from O_1 along the axis O_1O_λ'. Then the wave-front aberration of the wave-element considered is

$$W_\lambda' = (E_1A_1A_2 \ldots \ldots A_\lambda E_\lambda') - (P_0P_1P_2 \ldots P_\lambda P_{\lambda+1}), \qquad \ldots \ldots (1)$$

where the brackets denote optical path lengths.

A second element of wave proceeds from Q_1 along the ray $Q_1R_0P_0R_1R_2 \ldots \ldots$ $R_\lambda R_{\lambda+1}$, intersecting the Q_λ'-reference sphere in $R_{\lambda+1}$. If this element has a wave-front aberration V_λ', measured positively when the element is in advance of the Q_λ'-reference sphere, and such that this latter and the emergent wave coincide at $B_{\lambda+1}$—that is, V_λ' is zero for the element of wave which proceeds from Q_1 along the principal ray $E_1B_1B_1 \ldots \ldots B_\lambda B_{\lambda+1}$—then the wave-front aberration of the element considered is

$$V_\lambda' = (E_1B_1B_2 \ldots \ldots B_\lambda B_{\lambda+1}) - (R_0P_0R_1R_2 \ldots R_\lambda R_{\lambda+1}), \qquad \ldots \ldots (2)$$

where $B_{\lambda+1}$, $R_{\lambda+1}$ are the points of intersection of the emergent rays with the Q_λ'-reference sphere. Rewriting (2),

$$V_\lambda' = (E_1B_1B_2 \ldots \ldots B_\lambda B_{\lambda+1}) - (P_0R_1R_2 \ldots \ldots R_\lambda R_{\lambda+1}\bar{R}) + (R_{\lambda+1}\bar{R} - R_0P_0),$$

where $\bar{R}$ is the point of intersection of $R_\lambda R_{\lambda+1}$ with the O_λ'-reference sphere, it follows, using (1), that :

$$V_\lambda' = W_\lambda' + (R_{\lambda+1}\bar{R} - R_0P_0) + C_1, \qquad \ldots \ldots (3)$$

where

$$C_1 = \{(E_1B_1B_2 \ldots \ldots B_\lambda B_{\lambda+1}) - (E_1A_1A_2 \ldots \ldots A_\lambda E_\lambda')\}$$
$$- \{(P_0R_1R_2 \ldots \ldots R_\lambda R_{\lambda+1}\bar{R}) - (P_0P_1P_2 \ldots \ldots P_\lambda P_{\lambda+1})\}.$$

It can be shown that C_1 is zero (or negligible) for small angles of field.

From P_1 draw P_1h_1, P_1h_1' perpendicular to P_0P_1, P_1P_2, and meeting P_0R_1, R_1R_2 (produced if necessary) in h_1, h_1'. Then the following are differences in optical path length :

$$(P_0R_1 - P_0P_1) = N_1 . l_1 \sin I_1 + N_1 . D_0 (\sec x_1 - 1)$$
$$(R_1R_2 - P_1P_2) = N_2' . l_2' \sin I_2' - N_1' . l_1 \sin I_1' + N_2 . D_1(\sec x_2 - 1)$$
$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$
$$(R_\lambda\bar{R} - P_\lambda P_{\lambda+1}) = - N_\lambda' . l_\lambda \sin I_\lambda' + N_{\lambda+1} . D_\lambda (\sec x_{\lambda+1} - 1),$$

where $l_1 = P_1R_1$, $l_2 = P_2R_2$, $\ldots \ldots$ etc.; $x_1 = R_1\hat{P}_0P_1$, $\ldots \ldots$ etc.; $D_0 = P_0P_1$, $D_1 = P_1P_2$, $\ldots \ldots D_\lambda = P_\lambda P_{\lambda+1}$; N_1; $N_2 = N_1'$, $\ldots \ldots$ etc., are refractive indices ; and I, I' are the angles of incidence and emergence of the ray $P_0P_1P_2 \ldots \ldots P_\lambda$ at the successive surfaces of the system. Addition of these equations gives, the angles x being assumed small,

$$\{(P_0R_1R_2 \ldots \ldots R_\lambda R_{\lambda+1}\bar{R}) - (P_0P_1P_2 \ldots P_\lambda P_{\lambda+1})\}$$
$$= \tfrac{1}{2}x_1^2 . N_1D_0 + \tfrac{1}{2}x_2^2 . N_2D_1 + \ldots + \tfrac{1}{2}x_{\lambda+1}^2 . N_{\lambda+1} . D_\lambda, \qquad \ldots \ldots (4)$$

the other items cancelling because of the law of refraction between I and I'.

By a similar argument applied to the principal ray it can be shown that

$$\{(E_1B_1B_2 \ldots \ldots B_\lambda B_{\lambda+1}) - (E_1A_1A_2 \ldots \ldots A_\lambda E_\lambda')\}$$
$$= \tfrac{1}{2}y_1^2 . N_1d_0 + \tfrac{1}{2}y_2^2 . N_2d_1 + \ldots \ldots + \tfrac{1}{2}y_{\lambda+1}^2 . N_{\lambda+1}d_\lambda, \qquad \ldots \ldots (5)$$

where $d_0 = E_1A_1$, $d_1 = A_1A_2$, $\ldots \ldots$ etc.; and $y_1 = B_1E_1A_1$, and so on.

From (4) and (5)

$$C_1 = \tfrac{1}{2}y_1^2 . N_1(d_0 - D_0) + \tfrac{1}{2}y_2^2 . N_2(d_1 - D_1) + \ldots + \tfrac{1}{2}y_{\lambda+1}^2 . N_{\lambda+1}(d_\lambda - D_\lambda),$$

$$\ldots \ldots (6)$$

where a "paraxial" approximation has been made in putting $x_2 = y_2$, $x_3 = y_3$, ...etc., and where $x_1 = y_1 = \omega_1$; ω_1 being the angle of field $Q_1 \hat{E}_1 O_1$, which is small. If, now, retaining the paraxial approximation, there is written

$$y_1 = \frac{\omega_1}{V_1} . V_1, \quad y_2 = \frac{\omega_1}{V_1} . V_2, \ldots \ldots y_{\lambda+1} = \frac{\omega_1}{V_1} . V_{\lambda+1},$$

where $V_1, V_2 \ldots \ldots V_{\lambda+1}$ are angles from a traced paraxial principal ray, then

$$C_1 \doteqdot \tfrac{1}{2} \left(\frac{\omega_1}{V_1}\right)^2 \{N_1(d_0 - D_0) . V_1^2 + N_2(d_1 - D_1) . V_2^2 + \ldots$$

$$+ N_{\lambda+1}(d_\lambda - D_\lambda) . V_{\lambda+1}^2\} \quad \ldots \ldots (7)$$

which gives an approximate value for C_1. For small angles of field, powers of ω_1 higher than the first are neglected, and C_1 is regarded as zero. Equation (7) then provides an estimate of the most significant error arising from this approximation. On the other hand, given a "safe" limit for C_1, (7) provides an expression for ω_1, the size of field over which the approximation is sufficiently valid. C_1 is on an optical path length, and is one of the principal errors involved in V_λ', when C_1 is assumed zero in equation (3).

Attention has been confined so far to rays from Q_1, lying in the meridian section. It is, however, easy to show that the result (4), and therefore (7), hold equally for skew rays from Q_1. For, in this case, the elements of refracting surfaces at $P_1, P_2 \ldots \ldots P_\lambda$ are regarded as plane. Planes are constructed through each of these points perpendicular to the rays incident and emergent at the points. The differences in optical path length then become

$$(P_0 R_1 - P_0 P_1) = N_1 . l_1 \sin I_1 . \cos \phi_1 + N_1 . D_0 (\sec x_1 - 1)$$

$$(R_1 R_2 - P_1 P_2) = N_2 . l_2 \sin I_2 . \cos \phi_2 - N_1 . l_1 \sin I_1' . \cos \phi_1 + N_2 . D_1(\sec x_2 - 1)$$

and so on, where $\phi_1, \phi_2, \ldots \ldots$ etc., are the azimuths of the points of incidence $R_1, R_2, \ldots \ldots R_\lambda$. On addition, terms on the right-hand side again cancel, giving the result (4). Thus C_1 is zero for both meridian and skew rays, providing powers of ω_1 higher than the first are neglected. Furthermore, equation (7) still provides an estimate of the error involved in the approximation.

Construct, then, rectangular co-ordinates having O_1 as origin, $O_1 O_\lambda'$ as x-axis, and Q_1 in the XOZ-plane. Let $O_1 E_1 = r$, $Q_1 E_1 = p$; and let P_0 (on the incident wave from O_1) have co-ordinates (x, y, z). Then if Q_1 is the point (o, o, η), the length $Q_1 P_0$ is given by

$$Q_1 P_0^2 = x^2 + y^2 + (z - \eta)^2 = p^2 - 2z\eta.$$

In the diagram η_1 and ω_1 are shown negative, following the usual optical convention of signs.

Thus, since $O_1 \hat{E}_1 Q_1 = \omega_1$ is assumed small,

$$R_0 P_0 = Q_1 P_0 - p = -\frac{z . \eta}{p} - \tfrac{1}{2} \frac{z^2 . \eta^2}{p^3},$$

or, substituting $\dfrac{1}{p} = \dfrac{1}{r}\left(1 - \tfrac{1}{2}\dfrac{\eta^2}{r^2}\right)$,

$$R_0 P_0 = -\frac{z \cdot \eta}{r} - \tfrac{1}{2}\frac{z \cdot \eta^2}{r^2}\left[\frac{\eta}{r} - \frac{z}{r}\right].$$

Replacing z by polar co-ordinates (ρ_1, ϕ), where $\phi = 0$ is the positive direction of z, and ρ_1 is the perpendicular distance of P_0 from the axis, this becomes

$$R_0 P_0 = -\frac{\eta_1 \rho_1}{r_1} \cdot \cos\phi - \frac{\eta_1 \rho_1}{r_1} \cdot \cos\phi \cdot \tfrac{1}{2}\tan^2\omega_1\left(\tan\omega_1 - \frac{\rho_1 \cos\phi}{r_1}\right), \quad \ldots\ldots(8)$$

where the suffix 1 has also been given to η, ω and r.

Similarly if Q_λ' be joined to $P_{\lambda+1}$ (and continued), the intercept between the two reference spheres is very closely equal to $R_{\lambda+1}$; and, therefore,

$$R_{\lambda+1}\bar{R} = -\frac{\eta_\lambda' \rho_\lambda'}{r_\lambda'} \cdot \cos\phi - \frac{\eta_\lambda' \rho_\lambda'}{r_\lambda'}\cos\phi \cdot \tfrac{1}{2}\tan^2\omega_\lambda'\left(\tan\omega_\lambda' - \frac{\rho_\lambda' \cos\phi}{r_\lambda'}\right),$$
$$\ldots\ldots(9)$$

where ρ_λ' is measured at the exit pupil.

Subtraction of (8) and (9) gives for the optical path difference

$$(R_{\lambda+1}\bar{R} - R_0 P_0) = \left(\frac{N_1 \eta_1 \rho_1}{r'} - \frac{N_\lambda' \rho_\lambda' \rho_\lambda'}{r_\lambda'}\right)\cos\phi + C_2, \quad \ldots\ldots(10)$$

where C_2, a second-order term, is

$$C_2 = \frac{N_1 \eta_1 \rho_1}{r_1} \cdot \cos\phi \cdot \tfrac{1}{2}\tan^2\omega_1{}^2\left(\tan\omega_1 - \frac{\rho_1 \cos\phi}{r_1}\right)$$
$$- \frac{N_\lambda' \eta_\lambda' \rho_\lambda'}{r_\lambda'} \cdot \cos\phi \cdot \tfrac{1}{2}\tan^2\omega_\lambda'\left(\tan\omega_\lambda' - \frac{\rho_\lambda' \cos\phi}{r_\lambda'}\right).$$

It will be seen that C_2 is zero in the sagittal section, $\phi = \pi/2$. In the meridian section it has the value

$$C_2 = \frac{N_1 \eta_1 \rho_1}{r_1} \cdot \tfrac{1}{2}\tan^2\omega_1\left(\tan\omega_1 - \frac{\rho'}{r_1}\right)$$
$$- \frac{N_\lambda' \eta_\lambda' \rho_\lambda'}{r_\lambda'} \cdot \tfrac{1}{2}\tan^2\omega_\lambda'\left(\tan\omega_\lambda' - \frac{\rho_\lambda'}{r_\lambda'}\right). \quad \ldots\ldots(11)$$

This is zero if powers of ω_1 higher than the first are ignored. (11) then estimates the error involved in the approximation.

Substitution of (10) in (4) gives, ignoring C_1 and C_2,

$$V_\lambda' = W_\lambda' + \left[\frac{N_1 \eta_1 \rho_1}{r_1} - \frac{N_\lambda' \eta_\lambda' \rho_\lambda'}{r_\lambda'}\right] \cdot \cos\phi, \quad \ldots\ldots(12)$$

where r_1, r_λ' have signs such that, in figure 1, r_1 is positive and r_λ' negative. This will be so if

$$\begin{aligned} r_1 &= l_{\mathrm{pr}_1} - l_{\mathrm{parax}_1,} \\ r_\lambda' &= l_{\mathrm{pr}_\lambda} - l'_{\mathrm{parax}_\lambda,} \end{aligned}\right\} \quad \ldots\ldots(13)$$

with an obvious notation. Thus the wave-front aberration associated with the wave proceeding to Q_λ' consists of the spherical aberration W_λ' of the wave proceeding to O_λ', together with a coma-type term, which is zero in the sagittal

section ($\phi = \pi/2$) and has a maximum value in the meridian section ($\phi = 0$). The latter is merely reversed in sign in the section $\phi = \pi$. In the section $\phi = 0$, (12) becomes

$$V_\lambda' = W_\lambda' + \left[\frac{N_1 \eta_1 \rho_1}{l'_{\text{pr}_1} - l_{\text{parax}_1}} - \frac{N_\lambda' \eta_\lambda' \rho_\lambda'}{l'_{\text{pr}_\lambda} - l'_{\text{parax}_\lambda}} \right], \qquad \ldots \ldots (14)$$

and this can readily be given forms suited to special cases.

Thus, for telescopic systems in air ($N_1 = N_\lambda' = 1$),

$$V_\lambda' = W_\lambda' + [\rho_1 . \tan \omega_1 - \rho_\lambda' . \tan \omega_\lambda'], \qquad \ldots \ldots (15)$$

where ρ_1, ρ_λ' are measured in the (paraxially defined) entrance and exit pupils along the marginal ray. If h_1, h_λ' are the initial and final incidence heights of a paraxial ray trace, the magnification is

$$\frac{h_1}{h_\lambda'} = \frac{\tan \omega_\lambda'}{\tan \omega_1},$$

so that

$$V_\lambda' = W_\lambda' + \left[\rho_1 - \frac{h_1}{h_\lambda} . \rho_\lambda' \right] \tan \omega_1. \qquad \ldots \ldots (16)$$

Using the notation of figure 2, in which $\delta u_\lambda'$, the final angular aberration is
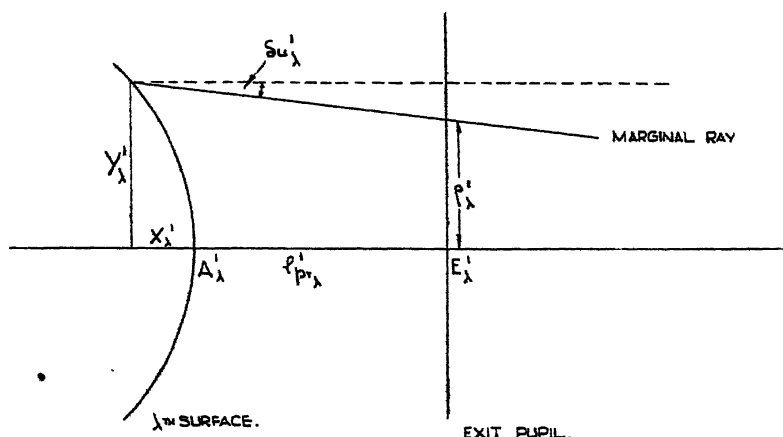
Figure 2.

shown as positive,

$$\rho_\lambda' = Y_\lambda' - (l'_{\text{pr}_\lambda} - X_\lambda') . \tan \delta u_\lambda'. \qquad \ldots \ldots (17)$$

From this it will be seen that, in the presence of spherical aberration ($\delta u_\lambda \neq 0$), ρ_λ', and hence the coma part of V_λ', will vary with l'_{pr_λ} the position of the exit pupil.

The case of telescopic and photographic objectives in air gives

$$V_\lambda' = W_\lambda' + \left[\rho_1 - \frac{F}{l'_{\text{pr}_\lambda} - l'_{\text{parax}_\lambda}} . \rho_\lambda' \right] \tan \omega_1, \qquad \ldots \ldots (18)$$

where use has been made of the relationship $\eta_\lambda' = F . \tan \omega_1$, F being the equivalent focal length of the system. In figure 3, u_λ' is the angle between the finally emergent marginal ray and the axis, $\delta u_\lambda'$, $\delta s_\lambda'$ are the angular and longitudinal

spherical aberration respectively of the marginal ray $P_{\lambda+1}O'_{m\lambda}$. Evidently, the angle $\delta u_\lambda'$ being small, and negative as shown,

$$\rho_\lambda' = r_\lambda' \cdot \sin(u_\lambda' - \delta u_\lambda')$$
$$= r_\lambda'\{\sin u_\lambda' + \delta s_\lambda' \cdot \sin u_\lambda' \cdot \cos u_\lambda' / r_\lambda'\}, \qquad \ldots\ldots(19)$$

since, with the given sign convention,

$$\delta u_\lambda' = -\frac{\delta s_\lambda' \cdot \sin u_\lambda'}{r_\lambda'}$$

Using this in (18) gives

$$V_\lambda' = W_\lambda' + \left[\rho_1 - F\sin U_\lambda'\left(1 + \frac{\delta s_\lambda' \cdot \cos U_\lambda'}{l'_{\text{pr}_\lambda} - l'_{\text{parax}_\lambda}}\right)\right] \cdot \tan \omega_1. \qquad \ldots\ldots(20)$$

It will again be seen that the coma part of V_λ' is independent of shifts of the diaphragm in the absence of spherical aberration ($\delta s_\lambda' = 0$) : on the other hand,
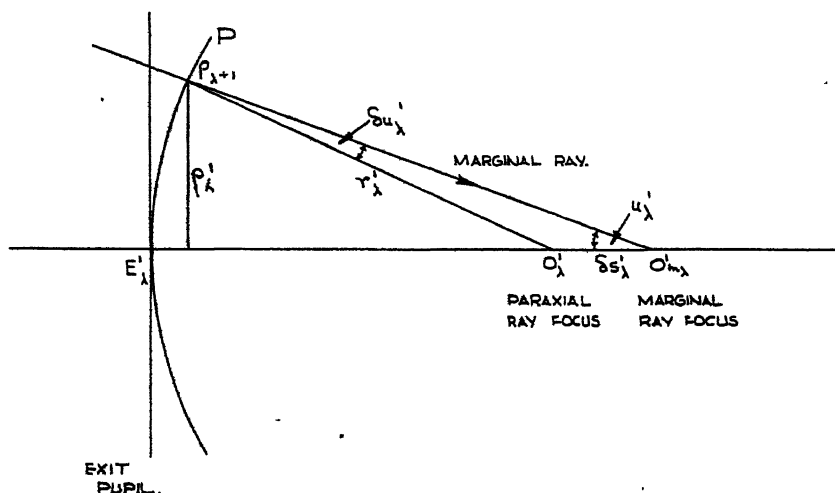
Figure 3.

V_λ' is independent of the spherical aberration when $l'_{\text{pr}_\lambda} = \infty$—that is, with a diaphragm in the anterior focal plane (telecentric stop). With the exit pupil at the posterior nodal point, $F = -(l'_{\text{pr}_\lambda} - l'_{\text{parax}_\lambda}) = -r_\lambda'$, and

$$V_\lambda' = W_\lambda' + [\rho_1 - (F - \delta s_\lambda' \cdot \cos U_\lambda') \cdot \sin U_\lambda']\tan\omega_1, \qquad \ldots\ldots(21)$$

or, since in this case $\omega_1 = \omega_\lambda'$,

$$V_\lambda' = W_\lambda' + \left[u'_{\text{parax}_\lambda} - \frac{F - \delta s_\lambda' \cdot \cos U_\lambda'}{F} \cdot \sin U_\lambda'\right]\eta_\lambda'. \qquad \ldots\ldots(22)$$

where $u'_{\text{parax}_\lambda}$ refers to a paraxial trace having the same initial incidence height as the marginal ray.

For systems in which both object and image distances are finite, (14) becomes

$$V_\lambda' = W_\lambda' + \left[N_1\eta_1\sin U_1 - N_\lambda'\eta_\lambda'\sin U_\lambda' \cdot \left(1 + \frac{\delta s_\lambda' \cdot \cos U_\lambda'}{l'_{\text{pr}_\lambda} - l'_{\text{parax}_\lambda}}\right)\right],$$
$$\ldots\ldots(23)$$

use having again been made of the result (19). Again it will be seen that V_λ' is independent of $\delta s_\lambda'$ when $l_{pr_\lambda}' = \infty$; and independent of l'_{pr_λ} when $\delta s_\lambda' = 0$. These results were obtained by Conrady (*loc. cit.* p. 375) using a "ray-inter-section" investigation. (23) can be put into a different form, thus

$$V_\lambda' = W_\lambda' + \left[\frac{\sin U_1}{u_1} - \frac{\sin U_\lambda'}{u_\lambda'}\left(1 + \frac{\delta s_\lambda' \cdot \cos U_\lambda'}{l'_{pr_\lambda} - l_{parax_\lambda}}\right)\right] H, \qquad \ldots\ldots(24)$$

where $H = N_1 u_1 \eta_1 = N_1' u'_1 \eta_1' = \ldots\ldots = N_\lambda' u_\lambda' \eta_\lambda'$ (the Lagrange Invariant) ; and small u's refer to paraxial angles. When $\delta s_\lambda' = 0$, or $l'_{pr_\lambda} = \infty$, this gives for the absence of coma the classical (but very restricted) result

$$\frac{\sin U_1}{u_1} = \frac{\sin U_\lambda'}{u_\lambda'}, \qquad \ldots\ldots(25)$$

The above formulae will give the coma part of the wave-front aberration even in the presence of large amounts of spherical aberration. The Rayleigh tolerance states simply that the total V_λ' must not exceed $\lambda/4$; Conrady's toler-ance applies only to the coma part of V_λ'. Given the same tolerance on W_λ', N_λ' might thus reach $\lambda/2$ at some parts of the aperture.

The coma-part of the expression for V_λ' includes all the coma terms depending on the first power of the field, i.e. those in which the wave-front aberration is pro-portional to the cube, fifth, seventh . . . . . powers of the aperture. The accuracy of the results can be easily checked by noting the values of C_1 and C_2. If, as in a good many telescope objectives, the incidence heights of the principal rays are all small, there can be written in (7)

$$V_2/V_1 = N_1/N_2$$

and so on, giving the useful result

$$C_1 = \tfrac{1}{2}N_1\omega_1^2[(d_0 - D_0) + (d_1 - D_1) + (d_2 - D_2) + \ldots\ldots + (d_\lambda - D_\lambda)]. \quad \ldots\ldots(26)$$

The correction term C_2, depending on powers of ω_1 higher than the square, will generally be smaller than C_1.

Although the above demonstrations have assumed the (more usual) system consisting of centred spherical refracting interfaces, it will be appreciated that essentially the results apply to other systems. Moreover, it will be seen that a similar analysis yields a " Sine-Condition " (based on an accurate principal ray-trace and another ray from the same object point) which will give the wave-front aberration associated with image points in the neighbourhood of an off-axis image point.

HERSCHEL'S CONDITION

By H. H. HOPKINS,

W. Watson and Sons, Barnet, Herts

ABSTRACT. The change in spherical aberration introduced by a change in object distance is discussed. Herschel's condition is shown to be valid for originally corrected conjugates. It follows as a special case of the general formulae derived.

———————

T HE satisfying of Herschel's condition by an optical system which is free from spherical aberration for given conjugates ensures freedom from spherical aberration for neighbouring pairs of conjugates. Attention does not seem to have been called to the invalidity of the usual form of the condition in the case when the system is not corrected for the original conjugates. It is, however, possible to formulate the condition in such a manner that it not only applies to such uncorrected systems, but such that the resulting generalized condition also measures directly the change in spherical aberration consequent upon a small change in the conjugates. The treatment resembles closely that which has been used elsewhere in connection with the Sine Condition (Hopkins, *Proc. Phys. Soc.* **58**, 92 (1946)).

In the figure, O_1 is an object point (real or virtual) situated on the axis $O_1 O_\lambda'$ of an optical system consisting of λ refracting surfaces. O_λ' is the Gaussian image of O_1 after λ refractions. Spheres ST, S'T' centred on O_1, O_λ' serve as reference surfaces, with respect to which the initially incident and finally emergent wave-fronts can be specified (*vide* Hopkins, 1945). These are conveniently chosen to intersect the axial points of the (paraxially defined) entrance and exit pupils. The latter are, respectively, Es and E_λ'.*

Assume the incident and emergent wave-fronts to coincide with the O_1- and O_λ'-reference spheres, respectively, at the points E_1, E_λ'. Let the element of wave from O_1 which is associated with the ray through P_0 be in advance of the O_1-reference sphere by an amount W_1, which is an optical path length. This

———————

* In the presence of spherical aberration the outer parts of the wave-front close up (or open out) more rapidly than the central parts. Because of this, the angular distribution of wave-front aberration depends upon the radii of the reference spheres. These are conventionally chosen, in each case, to refer the aberration to the entrance and exit pupils.

is, then, the initial wave-front aberration of the wave-element. Let W_λ' be the amount this element is in advance of the O_λ'-reference sphere upon emergence. This is, then, the final wave-front aberration of the wave-element. Since optical path lengths along rays between corresponding wave-fronts are equal,

$$W_\lambda' - W_1 + (P_0P_1P_2 \ldots\ldots P_1P_{\lambda+1}) = (E_1A_1A_2 \ldots A_\lambda E_\lambda'), \quad \ldots\ldots (1)$$

where the brackets denote optical path lengths.

Suppose, now, the object point be moved to Q_1, whose Gaussian image is Q_λ'. Let a wave-element, having initial wave-front aberration W_1, proceed from Q_1 along the ray $R_0P_0R_1R_2 \ldots R_\lambda R_{\lambda+1}$ (where R_0, $R_{\lambda+1}$ are on the spheres S_QT_Q, $S_Q'T_Q'$, which are centred on Q_1, Q_λ'. There can then be written

$$V_\lambda' - W_1 + (R_0P_0R_1R_2 \ldots\ldots R_\lambda R_{\lambda+1}) = (E_1A_1A_2 \ldots A_\lambda E_\lambda'), \quad \ldots\ldots (2)$$

where V_λ' is the final wave-front aberration of the wave-element on emergence from the system, measured with reference to S'_QT_Q', the Q_λ-reference sphere.

On subtraction (1) and (2) give

$$\begin{aligned} V_\lambda' - W_\lambda' &= (P_0P_1P_2 \ldots\ldots P_\lambda P_{\lambda+1}) - (R_0P_0R_1R_2 \ldots\ldots R_\lambda R_{\lambda+1}) \\ &= (P_0P_1P_2 \ldots\ldots P_\lambda P_{\lambda+1}) - (P_0R_1R_2 \ldots R_\lambda R_{\lambda+1}\bar{R}) \\ &\quad + (R_{\lambda+1}\bar{R} - R_0P_0). \end{aligned}$$

The first two brackets have been shown elsewhere to be equal, and they cancel, providing the angle $O_1R_0Q_1 = \Delta U_1$ is sufficiently small for powers of ΔU_1 higher than the first to be ignored. The error implied in this approximation has also been discussed elsewhere (Hopkins, *loc. cit.*). To this accuracy, the change in the wave-front aberration of the element considered is, therefore,

$$dW = (R_{\lambda+1}\bar{R} - R_0P_0). \quad \ldots\ldots (3)$$

Construct, then, plane rectangular co-ordinates having O_1 as origin, x-axis along the line O_1O_λ', and y-axis in the plane of the diagram and positive above the optical axis. Let P_0 have co-ordinates (x, y), and let E_1 be the point $(r_1, 0)$. Then

$$\begin{aligned} P_0Q_1^2 &= (x - \Delta x_1)^2 + y^2 \\ &= r_1^2 - 2x \cdot \Delta x_1, \end{aligned}$$

since P_0 is on the circle $x^2 + y^2 = r_1^2$. $\Delta x_1 = O_1Q_1$ is assumed small compared with r_1, and powers of $\Delta x_1/r_1$ higher than the first are therefore ignored. Taking square roots,

$$P_0Q_1 = r_1 - \frac{x \cdot \Delta x_1}{r_1}$$

to the required accuracy ; and further, ,

$$Q_1R_0 = Q_1E_1 = r_1 - \Delta x_1.$$

Thus

$$P_0Q_1 - Q_1R_0 = \Delta x_1\left(1 - \frac{x_1}{r_1}\right)$$

or

$$P_0R_0 = \Delta x_1 \cdot (1 - \cos\theta_1), \quad \ldots\ldots (4)$$

where θ_1 is the angle $P_0\hat{O}_1E_1$, O_1 being the paraxial object point.

If, now, O_λ' be joined to $P_{\lambda+1}$ (and continued), the intercept between the spheres $S'T''$, $S_Q'T_Q'$ is very closely equal to $R_{\lambda+1}\bar{R}$. Using, therefore, an expression for $R_{\lambda+1}\bar{R}$ similar to (4), the expression (3) becomes

$$dW = N_\lambda' . \Delta x_\lambda'(1 - \cos \theta_\lambda') - N_1 . \Delta x_1(1 - \cos \theta_1).$$

Thus, using the (paraxial) longitudinal magnification formula

$$N_\lambda' u_\lambda'^2 \Delta x_\lambda' = N_1 u_1^2 \Delta x_1 = H_L, \qquad \ldots \ldots (5)$$

the change in spherical aberration produced by the change of conjugates denoted by H_L is given by

$$dW = \left[\frac{1 - \cos \theta_\lambda'}{u_\lambda'^2} - \frac{1 - \cos \theta_1}{u_1^2}\right] . H_L, \qquad \ldots \ldots (6)$$

where u_λ', u_1 refer to paraxial angles associated with the original conjugates. This formula can be given forms suited to special cases.

Thus, in the absence of spherical aberration, $\theta_1 = U_1$, $\theta_\lambda' = U_\lambda'$, where U_1, U_λ' are angles from a marginal ray trace. (6) becomes

$$dW = \left[\left(\frac{\sin \frac{1}{2}U_\lambda'}{u_\lambda'/2}\right)^2 - \left(\frac{\sin \frac{1}{2}U_1}{u_1/2}\right)\right] . \frac{1}{2}H_L, \qquad \ldots \ldots (7)$$

which leads to the usual form of the Herschel condition. For dW will be zero, providing

$$\frac{\sin \frac{1}{2}U_\lambda'}{u_\lambda'/2} = \frac{\sin \frac{1}{2}U_1}{u_1/2}. \qquad \ldots \ldots (8)$$

If the system is uncorrected for the original conjugates, $\theta_\lambda' \neq U_\lambda'$, $\theta_1 \neq U_1$: (7), (8) are then no longer valid.

If, now,

$$\delta u_1 = -\frac{\delta s_1 . \sin U_1}{l_{\text{pr}_1} - l_{\text{parax}_1}} \quad , \quad \delta u_\lambda' = -\frac{\delta s_\lambda' . \sin U_\lambda'}{l'_{\text{pr}_\lambda} - l'_{\text{parax}_\lambda}}$$

are the initial and final angular spherical aberrations associated with a given ray, defined when the axial parts of the incident and emergent wave-fronts are at E_1, E_λ' respectively, and with the notation of an earlier paper (*loc. cit.*), then

$$\theta_1 = U_1 - \delta u_1, \qquad \theta_\lambda' = U_\lambda' - \delta u_\lambda'$$

and

$$\cos \theta_1 = \cos U_1 + \sin U_1 . \frac{\delta s_1 \sin U_1}{l_{\text{pr}_1} - l_{\text{parax}_1}}, \qquad \ldots \ldots (9)$$

$$\cos \theta_\lambda' = \cos U_\lambda' + \sin U_\lambda' . \frac{\delta s_\lambda' . \sin U_\lambda'}{l'_{\text{pr}_\lambda} - l'_{\text{parax}_\lambda}}$$

Substitution of these in (6) gives, after some reduction,

$$dW = \left[\left(\frac{\sin \frac{1}{2}U_\lambda'}{u_\lambda'/2}\right)^2 . \left(1 + \frac{1 - \cos U_\lambda'}{l'_{\text{pr}_\lambda} - l'_{\text{parax}_\lambda}} . \delta s_\lambda'\right)\right.$$

$$\left. - \left(\frac{\sin \frac{1}{2}U_1}{u_1/2}\right) . \left(1 + \frac{1 - \cos U_1}{l_{\text{pr}_1} - l_{\text{parax}_1}} . \delta s_1\right)\right]\frac{1}{2}H_L. \qquad \ldots \ldots (10)$$

In telescopic systems, the value of Δx is given in angular measure. Thus if h_1, h_λ are the paraxial incident heights at the first and last surfaces of any system and a change of conjugate Δu_1 results in a change $\Delta u_\lambda'$ in the object space, then generally

$$\Delta u = \frac{\Delta x \cdot u}{r} = \frac{Ax \cdot u^2}{h},$$

and, therefore, the longitudinal magnification formula becomes

$$N_\lambda' h_\lambda \Delta u_\lambda' = N_1 h_1 \Delta u_1 = H_\text{L} . \qquad \ldots \ldots (11)$$

It is this form which is used in telescopic systems. Let ρ_1, ρ_λ' be the intersection heights of a given ray in the paraxially defined entrance and exit pupils; then for small angles U_λ', U_1 (a " nearly " telescopic system), θ_λ', θ_1 are necessarily small, and

$$\theta_\lambda' = - \frac{\rho_\lambda'}{l'_{\text{pr}\lambda} - l'_{\text{parax}\lambda}}, \qquad u_\lambda' = \frac{h_\lambda}{l'_{\text{parax}\lambda}} \cdot$$

In the limit $l'_{\text{parax}\lambda} \to \infty$, the ratio

$$\frac{\theta_\lambda'}{u_\lambda'} = \frac{\rho_\lambda'}{h_\lambda} \cdot$$

Thus, for telescopic systems,

$$\frac{1 - \cos\theta_\lambda'}{u_\lambda'^2} = \frac{1}{2}\left(\frac{\rho_\lambda'}{h_\lambda}\right)^2, \qquad \ldots \ldots (12)$$

with a similar expression in the object space. Using (11), (12) in (6), the change of spherical aberration, consequent upon the change of conjugates H_L, is

$$dW = \left[\left(\frac{\rho_\lambda'}{h_\lambda}\right)^2 - \left(\frac{\rho_1}{h_1}\right)^2\right] \cdot \tfrac{1}{2} H_{L} \cdot \qquad \ldots \ldots (13)$$

It has been shown before (*loc. cit.*) that

$$\rho_\lambda' = Y_\lambda - (l'_{\text{pr}\lambda} - X_\lambda) \cdot \tan \delta u_\lambda',$$

with the notation of the earlier paper. In the absence of spherical aberration, Herschel's Condition for a telescope is thus

$$\frac{Y_\lambda}{h_\lambda} = \frac{Y_1}{h_1} \cdot \qquad \ldots \ldots (14)$$

It will be apparent that use of one or other of the forms of $\left(\dfrac{1 - \cos\theta}{u^2}\right)$ used in (10) and (13) will give results suited to photographic objectives (infinite object, finite image distance) and eyepieces (finite object, infinite image distance).

If it is assumed that for given conjugates a system is both free from spherical aberration and satisfies the sine-condition, there follow immediately results which have a bearing on a number of questions.

Thus, consider a corrected microscope objective of $\text{N.A.} = N_1 \sin U_1$, having a primary magnification m. A change of tube length $\Delta x_\lambda'$ gives, N_λ' being that of air $= 1$,

$$H_\text{L} = u_\lambda'^2 \cdot \Delta x_\lambda' = \left(\frac{N_1 u_1}{m}\right)^2 \cdot \Delta x_\lambda'.$$

The sine-condition is satisfied, so that

$$\frac{\sin \frac{1}{2} U_\lambda'}{u_\lambda'/2} \cdot \cos \frac{U_\lambda'}{2} = \frac{\sin \frac{1}{2} U_1}{u_1/2} \cdot \cos U_1/2,$$

or, the light coming from the short conjugate, u_λ' is small, and

$$\left(\frac{\sin \frac{1}{2} U_\lambda'}{u_\lambda'/2}\right)^2 - \left(\frac{\sin \frac{1}{2} U_1}{u_1/2}\right) = -\left(\frac{1 - \cos U_1}{u_{1}}\right)^2.$$

Substituting in (7) gives

$$d\dot{W} = -\left[\frac{N_1(1 - \cos U_1)}{m}\right]^2 \cdot \frac{1}{2}\Delta x_\lambda' . \qquad \ldots\ldots(15)$$

Thus for a $^1/_{12}$-in. objective having N.A. $= 1\cdot20$, $m = \times 100$, and using an oil for which $1\cdot53$,

$$dW = -0\cdot0000167 . \Delta x_\lambda',$$

i.e. a change of wave-front aberration of $\frac{1}{2}$ wave-length of sodium D light follows a change in tube-length of $17\cdot5$ mm.* With an objective of this type "good observers" claim to estimate tube-length to this order of accuracy, although the effects are generally unnoticeable at the "best" focus.

In a corrected telescopic system the sine-condition takes the form

$$\frac{Y_\lambda}{h_\lambda} = \frac{Y_1}{h_1},$$

which is identical with equation (14). Thus, in this case, Herschel's condition and the sine-condition are simultaneously satisfied.†

Consider a corrected photographic objective, satisfying the sine-condition. Let the long-conjugate change from ∞ to L_1, and let h_1 be the radius of the entrance pupil. Then by (11),

$$H_L = \frac{h_1^2}{L_1},$$

assuming $N_1 = 1$. The sine-condition gives

$$\frac{\rho_1}{h_1} = \frac{\sin \frac{1}{2} U_\lambda'}{u_\lambda'/2} \cdot \cos \frac{U_\lambda'}{2},$$

or, F being the equivalent focal length,

$$\left(\frac{\rho_1}{h_1}\right)^2 - \left(\frac{\sin \frac{1}{2} U_\lambda'}{u_\lambda'/2}\right)^2 = \left(\frac{1 - \cos U_\lambda'}{h_1}\right)^2 . F^2.$$

---

* Since writing this, Mr. A. G. Sabin has given me a similar value for a $\frac{1}{2}$-in. versalic objective obtained by accurate ray-tracing methods. For N.A.$=1\cdot16$, $m = \times 98$, a change in aberration of $1\cdot6$ wave-lengths was found to be produced by a change in tube-length of 50 mm. The above approximate formula gives $1\cdot3$ wave-lengths.

† This is also true at unit magnification for a copying system corrected for spherical aberration at that magnification.

The change in spherical aberration is, therefore,

$$dW = \tfrac{1}{2} \left(1 - \cos U_\lambda' \right)^2 \cdot \frac{F^2}{L_1} . \qquad \qquad \ldots\ldots(16)$$

Thus for an $F/1{\cdot}0$ objective,

$$dW = 0{\cdot}0056 \cdot \frac{F^2}{L_1},$$

and, for a 1-in. focus lens, a change in aberration of one wave-length occurs for $L_1 = 242$ inches. Further, for a given relative aperture, the value of L_1 for a given change in spherical aberration is proportional to the square of the equivalent focal length. For a given focal length it is approximately proportional to the fourth power of the relative aperture.

It will be seen that an identical analysis yields a "Herschel condition" applying to extra-axial object points. For first-order changes in the conjugates, the aberrational change is symmetrical about the principal ray (i.e. a "spherical aberration") and greater changes in the conjugates are required to introduce changes in the comatic and astigmatic aberrations.

THE THERMAL CONDUCTIVITY OF TEXTILES

By S. BAXTER,

Wool Industries Research Association, Leeds

ABSTRACT. A review is given of measurements of the thermal conductivity of natural clothing fibres and of formulae that have been proposed for the overall thermal conductivity of fibre-air mixtures in terms of their density and fibre conductivity.

Measurements of the conductivity of wool felts ranging in density from 0·01 to 1·0 gm./cc. have been made, using Hercus and Laby's method for determining the conductivity of air. The conductivity of solid keratin (horn) has also been measured by the Lees disc method, and it is suggested that wool fibres and horn have identical conductivities of $4{\cdot}62 \times 10^{-4}$ c.g.s. units at 0·7% regain and $5{\cdot}32 \times 10^{-4}$ c.g.s. units at 10·7% regain. The data given for wool felts and horn are used to obtain an empirical relation between conductivity and bulk density.

Results obtained with cotton fabrics indicate that published values of the thermal conductivity of cotton fibres are too high.

---

§1. INTRODUCTION

THE thermal conductivity of fabrics or masses of fibres is made up of two components: the heat transmitted through the air spaces and the heat transmitted through the textile fibres. Typical wool fabrics, as used for clothing purposes, have densities ranging from 0·1 to 0·3 gm. per cc., i.e. the air space in the fabrics varies from 92% to 77% of the volume of the fabric. Clearly this means that in a fabric the thermal conductivity of the air will be the controlling factor and the conductivity of the fibres will exert only a secondary influence; in particular the conductivity of a fabric of low bulk density will be practically independent of the fibres used. This result has been shown by many workers investigating the thermal insulation of fabrics.

If any differences in the thermal conductivity of various fibres are to be shown, the thermal conductivity measurements must be made on fabrics of high density, where the conductivity of the fibre becomes increasingly effective.

Few attempts have been made to determine the thermal conductivity of textile fibres from measurements of the conductivity of fibre-air mixtures. It was well known that low-density fibre-air mixtures had thermal conductivities approaching the value for air, and Speakman and Chamberlain (1930) have shown that the conductivity increases with the bulk density, depending on the type of fibre used.

The present work is an attempt to find an empirical formula for the wool conductivity-density curve from measurements of the conductivity of felts of increasing density and also of the conductivity of solid keratin. A satisfactory formula has been obtained, and the conductivity of any natural fibre can be predicted from its conductivity-density curve. Also, if the thermal conductivity of the fibre can be determined directly, as is the case with most synthetic fibres, then the conductivity of a fibre-air mixture of any density can be predicted.

§ 2. HISTORICAL

According to Schuhmeister (1877) the first investigations of the conducting power of cotton and wool fabrics were made by Peclet and reported by him in his *Handbook of the Science of Heat*. He concluded that the conducting power of various fabrics was little different from that of air. Later, Forbes (1872) found that the conducting power of cotton decreased with compression, a result which Schuhmeister interpreted as meaning that the conductivity of cotton was less than that of air. The present results show that it is quite possible, under certain conditions, for the conductivity of a fibre-air mixture to decrease as the density increases, although the fibre conductivity is greater than that of air. The problem of the conductivity of textile fabrics was first attacked seriously by Schuhmeister, using the method developed by Stefan for investigating the thermal conductivity of gases. In this method, the fibres are packed into the space between two concentric cylinders, the centre cylinder acting as an air thermometer. Schuhmeister showed that his results could be expressed by the formula

$$k_m = k_1 + aP, \qquad \ldots \ldots (1)$$

where k_m is the conducting power of the mixture, k_1 the conducting power of the gas, P the weight of textile fibre present, and a is a constant.

Taking the volume occupied by the fibres into account, the formula reduces to

$$k_m = k_1 + b\rho, \qquad \ldots \ldots (2)$$

where ρ is the bulk density of the fibres and b is a constant. For wool the value of the constant is 2·71, for silk 2·50 and for cotton 7·35. The range of values of bulk density used was small, the highest value being 0·15 gm./cc. In calculating the thermal conductivity of the fibres, Schuhmeister assumed that the fibres were uniformly distributed in all directions, or, for the purposes of calculation, that one-third were plane slabs parallel to the direction of heat flow and touching

both plates, and two-thirds were slabs at right angles to the direction of heat flow. On these assumptions,

$$k_m = \tfrac{1}{3}(k_1 v_1 + k_2 v_2) + \tfrac{2}{3}\left(\frac{k_1 k_2}{k_1 v_2 + k_2 v_1}\right), \qquad \ldots\ldots (3)$$

where k_1, k_2 and v_1, v_2 are the conductivity and fractional volume of air and of the fibres, so that

$$v_1 + v_2 = 1.$$

Using equation (3), and assuming the conductivity of air to be unity, Schuhmeister gave the value for cotton as 37, for wool 12 and for silk 11.

Rubner (1895), using the Stefan calorimeter and Schuhmeister's method of calculation, gave the conductivity of wool as 6·1, of silk 19·2 and of cotton and linen 29·9, taking the value for air as unity. Rubner measured the thermal conductivities of a wide range of keratin fibres and found no significant difference between them. The same result was also found for a wide range of cellulosic fibres, and for natural silks. His absolute values of conductivity are given in table 1, but it should be noted that they are not consistent with his ratios of the conductivities.

<div align="center">

Table 1. Rubner's values

Substance	Thermal conductivity k (c.g.s. units)
Air	$0\cdot532 \times 10^{-4}$
Mammal hairs	4·791
Silk	8·870
Plant fibres	14·199

</div>

Rubner also investigated the increase in the value of k_m with increasing density, but again he investigated only the region of very low densities, his maximum value being 0·12 gm./cc. His values for the constant in equation (2) were 3·65 for wool, 3·62 for silk and 5·9 for cotton.

One of the main reasons for the discrepancy between the results of Schuhmeister and Rubner was their choice of working in the region of low bulk density, where, as will be shown later, the conductivity of fibre-air mixtures behaves in a peculiar manner. One might have expected, however, that the agreement would have been better since Rubner's work appears to be an exact replica of Schuhmeister's. The error in estimating k_2 from equation (3) must be very large when low bulk densities are used. The higher the bulk density used to determine k_m, the smaller will be the error in estimating k_2. The validity of equation (3) is discussed later.

The theory of the conductivity of heterogeneous media is complex and has only been attempted for certain simple cases. Maxwell in his *Electricity and Magnetism* has worked out the conductivity of a medium formed by embedding spheres of conductivity k_2 in a medium of conductivity k_1, the spheres being a large distance apart. He also worked out the cases where the media were separated by planes both parallel and at right angles to the direction of heat flow, and the case of right prisms of rectangular cross-section, one dimension

being large compared with the other, and the length of the prisms being large compared with the larger cross-sectional dimension, placed either with their edges parallel or perpendicular to the direction of heat flow. These cases have limited application on account of the geometry implied. Lees (1900) has also worked out an expression for the conductivity of heterogeneous media from considerations of the heat flow through a medium formed by an equal number of equally long prisms of square cross-section having conductivities k_1 and k_2 and arranged in draught-board pattern, the planes of constant temperature being drawn through the diagonals of the cross-sections of the prisms. For this case Lees shows that the effective conductivity k_m is given by

$$k_m = k_1^{\frac{1}{2}} k_2^{\frac{1}{2}}. \qquad \ldots\ldots(4)$$

From this result he reasons that the thermal conductivity of a compound medium in the general case where the constituents are in the form of long prisms with their axes perpendicular to the direction of heat flow is given by

$$k_m = k_1^{v_1} k_2^{v_2}, \qquad \ldots\ldots(5)$$

where v_1 and v_2 are the fractional volumes of the media of thermal conductivity k_1 and k_2 respectively. It must be noted, however, that, even in the general case, equation (5) only holds when the prisms are square in cross-section and are arranged in some form of draught-board pattern with their edges touching along their length.

§3. EXPERIMENTS WITH THE LEES DISC

The experimental method first used by the author to determine the absolute conductivity of felts and fabrics was that of the Lees disc (1898). The apparatus is sufficiently well known only to require brief description here. Its essentials, shown in figure 1, are the brass plates C and U, which are screwed together with a
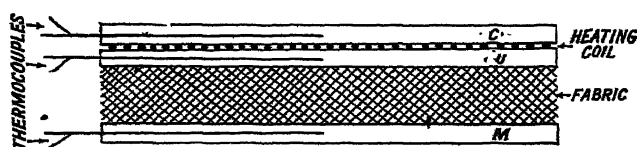
Figure 1. Lees' disc apparatus.

heating coil between them. The fabric to be tested is placed between this hot plate and a lower plate M, and the whole block is then suspended in an enclosure of constant temperature until the readings of heater current and voltage and the plate temperatures, which are measured by thermocouples, become steady. The calculation of the results is somewhat tedious, and for details the original paper should be consulted.

The conductivity of layers of thin cotton, wool and rayon fabrics were measured by this method, and in each case it was found that the thermal conductivity increased with the number of layers used. Figure 2 shows the values obtained using 1, 2, 4 and 8 layers of cotton fabric, both with the heater above and below the fabric. The results can be explained by assuming a constant air layer of thickness $d_a/2$ at each plate, the rest of the space being assumed to

have a constant thermal conductivity k, figure 3. The measured conductivity K will be a function of the thickness d, and will be given by

$$K = \frac{kk_a d}{kd_a + k_a(d - d_a)}. \qquad \ldots\ldots(6)$$

From equation (6) it is seen that $K \to k$ as $d \to \infty$. Curve (1) of figure 2 is for equation (6) with $k_a = 0.6 \times 10^{-4}$ c.g.s.; $k = 2.72 \times 10^{-4}$ and $d_a = 1.1 \times 10^{-2}$ cm. Curve (2) of figure 2 is for $k = 2.2 \times 10^{-4}$ and $d_a = 1.1 \times 10^{-2}$ cm. Although the
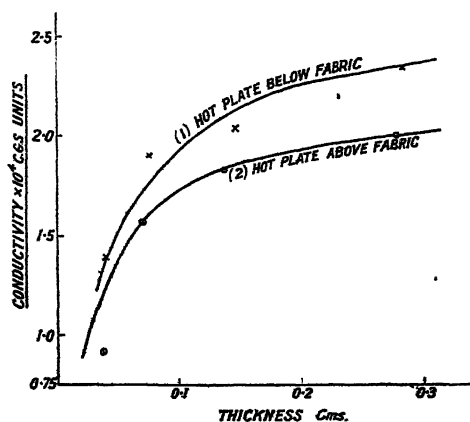
Figure 2. Conductivity of layers of cotton fabric as measured by Lees' disc apparatus.

curves fit the experimental points, the large value of d_a in comparison with the single fabric thickness 0·04 cm. gave rise to doubts as to their validity. In Lees' calculation the edges of the material tested were assumed to have the same emissivity as the metal plates. This equality can be achieved with a solid sample merely by varnishing the whole block after assembly with, say, shellac, but when using textiles such a course is impossible. Also, when working with textiles, heat can be lost from the edges of the fabric by convection of the air; i.e. by hot air leaving the centre of the fabric and diffusing out at the edges. Such
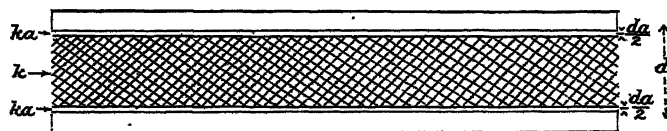
Figure 3. Boundary resistances.

effects cannot easily be allowed for in the calculations, and almost certainly contribute to the results of figure 2. In view of these difficulties, the use of the Lees disc to measure the conductivity of textile-air mixtures was abandoned.

An attempt was made to estimate the thermal conductivity of textile fibres by placing them in liquids and measuring the thermal conductivity of the mixture using Lees' disc. If this were done using liquids which had both higher and lower conductivities than the fibres, it was hoped that the conductivity of the fibres could be estimated. The results obtained using cotton and wool were

somewhat erratic, depending on whether or not the liquid was absorbed by the fibres, but the value for the conductivity of wool appeared to be 3.5×10^{-4} c.g.s. units and for cotton 6.0×10^{-4} units. No great accuracy is claimed for these results.

§4. EXPERIMENTS WITH THE APPARATUS OF HERCUS AND LABY

As mentioned earlier, textile fabrics contain a very high proportion of air by volume, and in measuring the thermal conductivity of a fabric, it would appear better to use an apparatus designed to measure the thermal conductivity of gases rather than one designed to measure the thermal conductivity of solids. The types of apparatus used to measure the conductivity of gases fall generally into two classes : the hot-wire type, in which the heat losses from a heated wire in a cylindrical tube are measured, and the parallel-plate method, where the heat transmitted through the material from one plate to a colder plate is measured. An apparatus of the second type was most suitable for our purpose, and Hercus and Laby's (1919) design has been used. The apparatus, which is shown in figure 4, was made in two halves, one unit comprising the top plate, centre plate
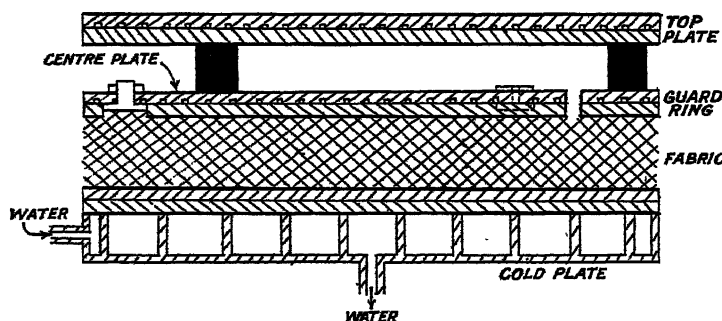
Figure 4.   Apparatus based on that of Hercus and Laby.

and guard ring, the other unit being the cold plate.   Each plate was made of two sections which were screwed together.   The top sections in the case of the top plate, centre plate and guard ring had spiral grooves cut into their lower faces, the heating coils being wound into these grooves.   The bottom section of each of the three plates had a radial groove cut in its upper face to take the thermo-couples.   The top section of the bottom plate also had a radial groove cut into its underside for thermocouples.   The cold plate was kept at a constant tem-perature by circulating cold water (26° c.) through it from a thermostated tank.   The apparatus could be used with the hot plate above the fabric and the cold plate below or vice versa.   Full details of the experimental technique and method of calculating the results will be found in Hercus and Laby's paper (1919).   Briefly the method is to heat the guard ring and top plate to a temperature slightly higher than the centre plate, so that all the heat input into the centre plate is transmitted throught the material to the cold plate.   Corrections are applied for the small amount of heat gained from the top plate and the guard ring.   All the outer surfaces of the plates were silver-plated, and when radiation corrections

were applied, the emissivity of the plates was taken as $4 \cdot 1 \times 10^{-12}$ cal./sec./cm.^{2}/deg.^{4} The e.m.f. generated by the thermocouples, which had previously been calibrated, was measured on a Cambridge potentiometer.

(a) Results using air

As a check on the apparatus, the thermal conductivity of air was determined. The air was room air. The results obtained were reduced to k_0, the value of the conductivity at $0°$ c. by use of the formula, $k_t = k_0(1 + 0 \cdot 003\,t)$, as used by Hercus and Sutherland (1934). The values obtained are given in table 2.

The values obtained with the heater below the air gap confirm the well-known result that convection plays no appreciable part until the thickness of the air gap becomes greater than about 1 cm., when the effect becomes quite large, as is seen by the last value of the table. The mean value obtained for the conductivity of air may be compared with some of the most accurate measurements made to date. Kannuluick and Martin (1934), using the hot-wire method, give k_0 as $5 \cdot 76 \times 10^{-5}$ c.g.s. units, and Hercus and Sutherland (1934), using the parallel-plate method of Hercus and Laby, give k_0 as $5 \cdot 72 \times 10^{-5}$ c.g.s. units. The agreement of the present results with these values serves as a check on the calibrations of the thermocouples and electrical measuring instruments and shows that no systematic error was present.

Table 2

Hot plate above		Hot plate below	
Thickness of layer (cm.)	k_0 (c.g.s. units)	Thickness of layer (cm.)	k_0 (c.g.s. units)
0·150	5·44 × 10⁻⁵	0·171	5·58 × 10⁻⁵
0·153	5·64	0·184	5·64
0·450	5·61	0·476	5·84
0·458	5·92	0·966	5·83
0·953	5·82		
1·946	6·09	1·968	12·48 *

Mean 5·75   Mean 5·72

* Not taken into account in determining the mean value.

(b) Results using fabrics

The thermal conductivity of layers of thin fabrics made from wool, cotton and viscose rayon were determined. The wool and cotton cloths were made as nearly identical as possible, but the rayon cloth was thinner than the others. The wool and cotton cloths had weights of 4·3 oz. per sq. yd. and the rayon cloth weighed 3·1 oz. per sq. yd., the thicknesses under a pressure of 5 gm./cm.² being 0·037 cm. for the wool fabric, 0·035 cm. for the cotton fabric and 0·023 cm. for the rayon fabric. The measurements of conductivity were done with the fabrics at room regain,* and with the hot plate both above and below the fabric,

* Regain is the moisture content expressed as a percentage of the dry weight.

* *S. Baxter*

Table 3

Number of layers	Bulk density (gm./cc.)	Thickness (cm.)	$10^4 \times$ conductivity (cal./cm./sec./°c.)	Mean temp. (c.)
		Heater above fabric : Cotton		
8	0·450	0·252	1·70	27·9
8	0·450	0·252	1·69	28·0
17	0·467	0·515	1·74	32·1
	Mean 0·455		Mean 1·71	Mean 29·3
		Wool		
9	0·475	0·310	1·25	29·3
18	0·480	0·612	1·26	32·3
	Mean 0·477		Mean 1·26	Mean 30·8
		Rayon		
12	0·520	0·229	1·45	28·7
16	0·515	0·310	1·45	28·9
16	0·515	0·310	1·48	27·8
	Mean 0·517		Mean 1·46	Mean 27·5
		All-wool greatcoat material		
1	0·295	0·241	1·05	28·1
1	0·295	0·241	1·02	29·1
2	0·285	0·501	1·07	29·4
2	0·285	0·501	1·07	29·5
3	0·287	0·740	1·09	31·6
4	0·287	0·986	1·13	32·8
5	0·296	1·201	1·12	33·6
	Mean 0·290		Mean 1·08	Mean 30·6
		Heater below fabric : Cotton		
9	0·460	0·276	1·62	29·4
17	0·460	0·521	1·71	32·6
17	0·460	0·521	1·65	27·5
	Mean 0·460		Mean 1·66	Mean 29·8
		Wool		
9	0·482	0·307	1·17	29·2
9	0·482	0·307	1·18	27·1
18	0·468	0·630	1·26	32·0
18	0·468	0·630	1·28	31·7
	Mean 0·475		Mean 1·22	Mean 30·0
		Rayon		
10	0·531	0·189	1·35	29·0
15	0·524	0·287	1·40	30·1
	Mean 0·527		Mean 1·38	Mean 29·6
		All-wool greatcoat material		
3	0·290	0·735	1·08	33·8
5	0·287	1·239	1·09	33·3
	Mean 0·289		Mean 1·09	Mean 33·6

in order to determine whether or not convection losses through the fabric were possible. The values obtained are given in table 3.

The results show that there is little change in conductivity with increasing thickness, and that in most cases where an increase is shown, the higher value has been obtained at a higher temperature. The values obtained with the heated plate below the fabrics show that no convection took place through the fabrics. The bulk densities of the thin fabrics when in the apparatus and at room regain were very similar ; the mean values expressed as gm. of fibres per cc. were 0·476 for the wool fabric, 0·458 for the cotton fabric and 0·522 for the rayon fabric. Since the weaves of the fabrics were similar, it must be concluded that the wool fibre has the lowest conductivity of the three materials investigated.

(c) *Variation of conductivity with density*

In order to be able to make an estimate of the conductivity of the wool fibre, it is necessary to determine the curve of conductivity against density for wool at a constant regain and then to extrapolate the curve to find the conductivity corresponding to a density of 1·3 gm./cc. In order to do this, a length of felt was specially prepared with a density of approximately 0·1 gm./cc. and thickness
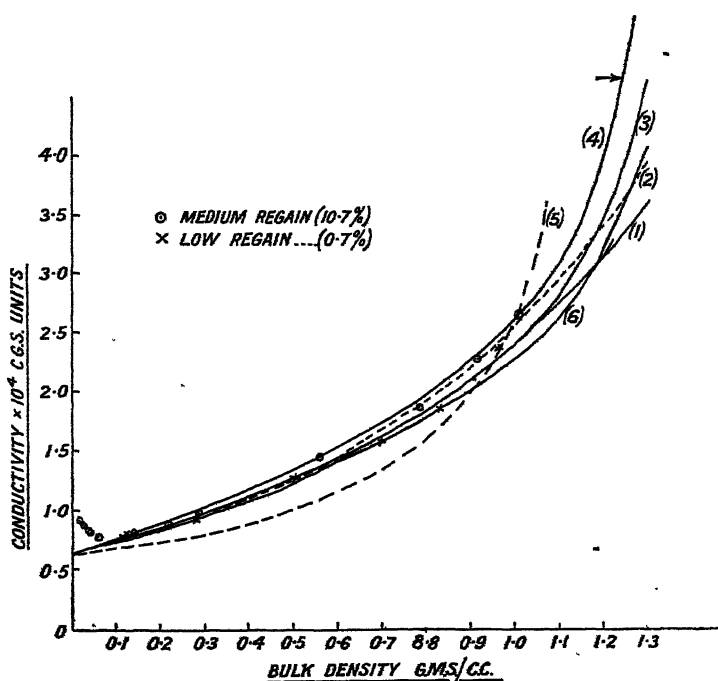
Figure 5. Conductivity against bulk density for wool.

approximately ½ inch. Values at higher densities were obtained by compressing layers of the felt together at a temperature of 100° c. and allowing them to cool before releasing the compression. If the wool had a regain not less than about 20%, it was found that the heat and compression treatment caused a permanent set. Excess water was not necessary, as the saturation regain of wool at 100° c.

is in the region of 20%. Using this technique, densities up to 1·0 gm./cc. were obtained. For densities less than 0·1 gm./cc. the felt was carded dry and placed in laps between the plates, which were separated by spacing blocks of ebonite between the guard ring and the cold plate. The conductivity of each felt of density greater than 0·1 gm./cc. was measured at a low (0·7%) and a medium (10·7%) regain, the regains being based on the dry weight obtained by placing the felt in an oven maintained at 100° c. The values obtained are shown graphically in figure 5.

(d) *Variation of conductivity with regain*

The effect of increasing the regain is to increase the bulk density and also the conductivity, the magnitude of the effect depending on the bulk density. At low densities the regain effect can hardly be detected owing to the small part played by the fibres and also as at low temperatures the conductivity of air is negligibly altered by increasing the relative humidity. This has been verified in the present investigations, which are in line with the results of Gruss and Schmick (1928), who showed that introducing 5% of water vapour into the air caused the thermal conductivity of air to increase by 1%.
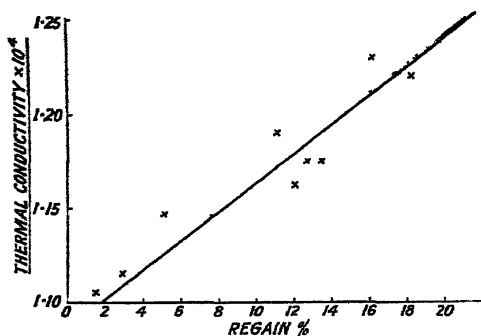
Figure 6.   Effect of regain on conductivity of wool felt of density 0·335 gm./cc.

The effect of regain on a felt of density 0·30 gm. per cc. when dry was investigated. The density of the felt varied slightly with regain, and the values of the conductivity were corrected, using the curve of figure 5 to correspond in each case to a density of 0·335 gm./cc. The values are shown in figure 6, and the conductivity is seen to increase linearly with increasing regain. The slope of the line is $7·7 \times 10^{-3}$ units of conductivity per unit change in regain. This value will of course depend on the bulk density of the material. The experimental points are all within 2% of the straight line.

§5. THE CONDUCTIVITY OF HORN

Since both wool and horn are keratin, the estimated values for the conductivity of wool would be expected to be similar to the conductivity of horn. The thermal conductivity of horn, in a direction at right angles to the axis of the long chain molecules, was determined at various regains by means of the Lees disc apparatus. A piece of horn was flattened by boiling whilst clamped in a flat position, and then

turned on a lathe into a disc of the same diameter as the Lees discs, and of thickness 0·226 cm. when dry. The horn warps badly as the regain is altered, and the specimen was held clamped flat between the plates of the apparatus, paraffin wax being used to ensure good thermal contact between the specimen and the plates. The conductivity of the horn was investigated over a wide range of regains, and the values are given in figure 7.

The thermal conductivity of mixtures can in many cases be expressed by a formula of the type

$$k_m = k_0(1 + a'C_m),$$

where k_m is the conductivity of the mixture, k_0 the conductivity of the pure substance, C_m the concentration of the added substance, and a a constant. Water molecules are absorbed by wool in two states, some being absorbed on to localized
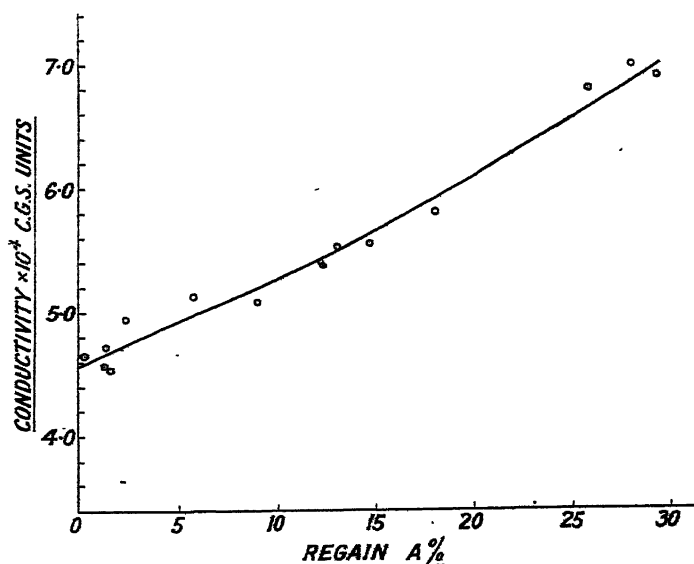
Figure 7. Effect of regain on the conductivity of horn.

sites and others existing in the liquid state. The proportion of water in the two states at any regain has been worked out by Cassie (1945). Assuming these two types of water, the conductivity of horn may be expressed by the equation

$$k_m = k_0(1 + \alpha C_X + \beta C_{(A-X)}), \qquad \ldots\ldots(8)$$

where k_0 is the conductivity of the dry horn, k_m that of the horn when at regain m ; C_X is the concentration of water absorbed on the localized sites, and $C_{(A-X)}$ is the concentration of water in the liquid state, whilst α and β are constants.

The values of C_A, C_X and $C_{(A-X)}$ for various regain values have been evaluated from table 3 of Cassie's paper, and are given in table 4.

The curve in figure 7 is for equation (8), the constants being obtained by making the curve to pass through the points $k_m = 5\cdot3 \times 10^{-4}$, $A = 10\cdot8\%$ and $k_m = 6\cdot6 \times 10^{-4}$, $A = 25\cdot2\%$. The values of the constants are $\alpha = 49\cdot4$ and $\beta = 133$, indicating that the free water contributes more to the thermal conductivity than the water found on localized sites.

8-2

Table 4

Regain (A)	C_A (gm./cc.)	C_X (gm./cc.)	$C_{(A-X)}$ (gm./cc.)
%			
3·6	0·045	0·041	0·004
7·2	0·086	0·082	0·004
10·8	0·125	0·113	0·012
14·4	0·165	0·141	0·024
18·0	0·200	0·159	0·041
21·6	0·232	0·169	0·063
25·2	0·260	0·174	0·086
28·8	0·286	0·174	0·112
32·5	0·314	0·174	0·140

§6. INTERPRETATION OF RESULTS

There are two possible equations which may be applied to the experimental points of figure 5 ; they are the Lees formula, equation (5), and Schuhmeister's formula, equation (3). The orientation of the wool fibres in the compressed felts used do not agree with the conditions assumed in deducing either of the two formulae, and any equation which fits the data must be taken as a purely empirical equation without any real theoretical foundation. Curve (1) of figure 5 is for Lees' equation (5), with $k_1 = 0.63 \times 10^{-4}$, the conductivity of air at the temperature of the experiments, $k_2 = 3.63 \times 10^{-4}$, and curve (2) is for equation (5), with $k_1 = 0.63 \times 10^{-4}$ and $k_2 = 3.92 \times 10^{-4}$. The two curves fit the experimental points quite well, except for densities less than 0·2 gm./cc., and predict the conductivity of wool at the lower regain (0·7%) as 3.63×10^{-4}, and at the higher regain (10·7%) as 3.95×10^{-4}. The values of the conductivity of horn at the same two regains are 4.62×10^{-4} and 5.32×10^{-4}.

Schuhmeister's formula, equation (3), may be written

$$k_m = x(k_1 v_1 + k_2 v_2) + y \frac{k_1 k_2}{k_1 v_2 + k_2 v_1}, \qquad \ldots \ldots (9)$$

where $x + y = 1$.

If x is unity, we have k_m increasing linearly with density from k_1 to k_2, and if y is unity we have a hyperbolic relation with k_m, increasing slowly with density at first and then more rapidly as the density increases. Equation (9) has been used with $y = 1$ for water-vapour diffusion by Peirce, Rees and Ogden (1945), but this by no means fits the experimental points, as is seen by curve (5) of figure 5, which is equation (9) for $y = 1$, $k_2 = \infty$ and the density of the fibres 1·3 gm./cc. As there are no means of determining directly the values of x and y, the only method is to take the value of k_2 as determined for horn and then to fit the equation to the experimental points and so determine x and y. This method gives the values of x and y as 0·29 and 0·71 respectively. Curve 3 of figure 5 is for equation (6) with $x = 0.29$, $y = 0.71$, $k_1 = 0.63 \times 10^{-4}$, $k_2 = 4.62 \times 10^{-4}$, and curve 4 is with $k_2 = 5.32 \times 10^{-4}$. The curves again fit the experimental points within experimental error. If we take $x = \frac{1}{3}$ and $y = \frac{2}{3}$, which are Schuhmeister's values, and fit the equation to the low regain experimental points, $k_m = 1.425 \times 10^{-4}$ c.g.s. and $\rho = 0.6$ gm./cc., we find that k_2 is 4.04×10^{-4} c.g.s. units, and with this value

we get curve (6). This curve is almost identical with curves (1) and (2) over the region explored experimentally.

We have, therefore, to decide which equation, type (3) or type (5), should be used. Both equations give curves which fit the experimental points equally well over the range measured, but the Schuhmeister formula, with $x = 0.29$ and $y = 0.71$, takes in the values found for the conductivity of horn and, therefore, may be considered to give a better empirical formula than the Lees equation or Schuhmeister's equation with $x = \frac{1}{3}$, $y = \frac{2}{3}$.

At densities less than 0·1 gm./cc. neither formula fits the experimental points. The apparent increase in conductivity as the density falls below 0·1 gm./cc. is probably due to convection. The temperature gradient along a fibre will be different from the temperature gradient in the air adjacent to it. This will give convection currents whose magnitude will depend on the fibre surface area and the average distance between the fibres. As the density increases, the surface area increases, but the mean distance between the fibres decreases and a maximum would be expected in the (conductivity v. density) curve. This is seen to be the case: the convection effect rises to a maximum at a density of 0·01 gm./cc. and then

Table 5

Substance	Density (gm./cm^3)	$10^4 \times$ thermal conductivity (c.g.s. units)
Polyvinyl chloride–acetate	1·34–1·37	3·95
Polyvinyl chloride	1·2 –1·6	3·9–4·0
Cellulose acetate (transparent)	1·27–1·32	5·4
Cellulose acetate (pigmented)	1·37–1·56	5·3–8·7
Ethyl cellulose	1·10–1·25	5·6

decreases as the density increases, and becomes negligible at densities greater than 0·2 gm./cc.

Equation (9) with $x = 0.21$ and $y = 0.79$, may be used to give estimates of the conductivities of the fibres for the three thin fabrics of table 3. Using the mean values of the density and fabric conductivity, the conductivity of the wool fibre is found to be 4.6×10^{-4}, of the cotton fibres 11.0×10^{-4}, and the viscose rayon fibres 6.9×10^{-4}, the regains being approximately 10, 5 and 10% respectively. Taking the conductivity of air, 0.63×10^{-4}, as unity, the relative conductivity of wool becomes 7·3, of cotton 17·5 and of viscose rayon 11·0. The value for wool lies between the estimates of Rubner and Schuhmeister, and the value for cotton is roughly half their values. The value obtained for the viscose rayon compares favourably with those given for cellulose derivatives in table 5 below, although the conductivity of viscose—regenerated cellulose—is not known. The result for cotton, taken together with the value indicated by the experiments with liquids, suggests that Schuhmeister and Rubner's values are too high.

The thermal conductivity of synthetic fibres can be measured directly, as the materials are readily available in sheet form, and, knowing the fibre conductivity, the conductivity of a fabric can be estimated from its density and equation (9). The thermal conductivities of many of the synthetic materials used as textiles are not available, but a few are given in table 5 (cf. Simonds and Ellis, 1943).

No mention is made of the regain of the material at which the measurements were made.   Wool is seen to have a conductivity comparable with that of the artificial materials, whilst cotton appears to have a considerably higher conductivity.

ACKNOWLEDGMENTS

REFERENCES

CASSIE, 1945.   *Trans. Faraday Soc.* **41,** 458.
FORBES, 1872.   *Proc. Roy. Soc. Edin.* **18,** 66.
GRUSS and SCHMICK, 1928.   *Wiss. Veröff. Siemens-Konz.* **7,** 202.
HERCUS and LABY, 1919.   *Proc. Roy. Soc.* A, **95,** 190.
HERCUS and SUTHERLAND, 1934.   *Proc. Roy. Soc.* A, **145,** 599.
KANNULUICK and MARTIN, 1934.   *Proc. Roy. Soc.* A, **144,** 496.
LEES, 1898.   *Philos. Trans.* **191,** 399.
LEES, 1900.   *Phil. Mag.* **49,** 221.
PEIRCE, REES and OGDEN, 1945.   *J. Textile Inst.* T. 169.
RUBNER, 1895.   *Arch. Hyg.* **24,** 265.
SCHUHMEISTER, 1877.   *Ber. K. Akad. Wien (Math.-Naturw. Klasse),* **76,** 283.
SIMONDS and ELLIS, 1943.   *Handbook of Plastics* (Chapman and Hall), pp. 40 and 48.
SPEAKMAN and CHAMBERLAIN, 1930.   *J. Textile Inst.* T. 29.

THE DEPENDENCE OF PRESSURE-VARIATION, WITH DEPTH IN A LIQUID, ON THE VERTICAL ACCELERATION OF THE LIQUID

BY KERR GRANT,

University of Adelaide

THE experimental demonstration here illustrated was suggested to me by the obvious scepticism of a class of first-year students to a remark that water in a falling vessel would not issue from an orifice in the bottom of the vessel.

To demonstrate the truth of the statement, a flask filled with coloured water was attached to a spiral spring about 3 feet in length, which permitted of an oscillation of the flask in a vertical direction (*v.* figure 1).   To the bottom of the flask was sealed a short tube, the opening of which was drawn down to a capillary of some 2 mm. diameter.   The period of oscillation of the flask was about 2 sec. and an amplitude of 6 inches could be given to it.   Thus it was subjected to a vertical acceleration ranging sinusoidally from about 5 ft. sec. sec. upwards to the same value downwards, and the value of gravity thus effectively varied from $(32+5)=37$ to $(32-5)=27$ ft. sec./sec.

This variation, producing, of course, a proportional variation in the pressure

(above atmospheric) at the orifice, was sufficient to cause alternately cessation and occurrence of flow of the liquid from the orifice, as shown in figure 2.

Figure 1. Flask containing inky water suspended on spiral spring and executing vertical oscillations of period about 3 seconds.   Beneath flask is a rotating table covered with white paper.

Figure 2. Shows intermittent character of flow of fluid from opening of fine tube. Due to capillary effect at the orifice, the interval during which flow is inhibited is longer than that during which it occurs.

THE HYSTERESIS OF THE ANGLE OF CONTACT OF MERCURY

By The Rev. G. D. YARNOLD,
University College, Nottingham

MS. received 19 *July* 1945

ABSTRACT. A dynamical method is described for the observation of the advancing and receding angles of contact of mercury at very low velocities. It is suggested that the irreversible part of the work expended when the line of contact of the liquid traverses the solid surface is different for the two directions of motion.

§ 1. INTRODUCTION

THE angle of contact of mercury differs widely according as the liquid is advancing or receding over the surface of a solid. It has already been shown (Yarnold, 1938, 1940) that the difference between the cosines of the advancing and receding angles of contact against walls of glass, Pyrex and quartz is independent of the velocity of the line of contact throughout the range from 3 to 30 cm./sec. The hysteresis of the angle of contact at very low velocities has now been studied by measurement of the force exerted by a rising or falling mercury surface on a suspended sphere which is partly immersed in the liquid. An expression for the force experienced by the sphere may be derived as follows.
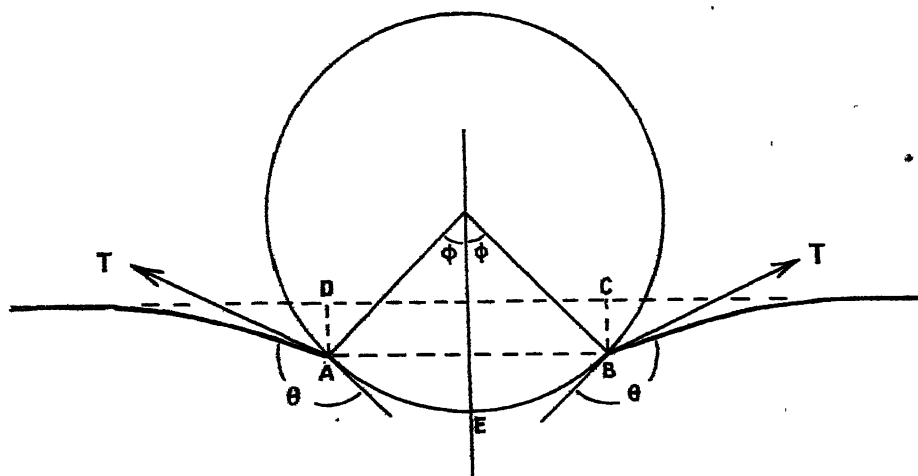
Figure 1.

Consider a sphere of radius a whose centre is at height y above the plane surface of the mercury. Whether the mercury surface is elevated or depressed in the neighbourhood of the sphere depends upon the value of y and upon the motion of the sphere. In figure 1 the mercury surface is represented as depressed in the region of the sphere, with which it makes an angle of contact θ. The lines joining the circle of contact with the centre of the sphere make an angle ϕ with the vertical. The upthrust of the liquid may be divided for convenience into two parts : that due to the displacement of the segment ABE and that due to the dis-

placement of the cylinder ABCD. The forces due to surface tension act in directions inclined at an angle $(\theta + \phi - \pi)$ to the horizontal. The total upward force on the sphere is therefore given by the equation

$$U = \pi a^3 \rho g \left(\frac{2}{3} - \cos \phi + \frac{1}{3} \cos^3 \phi \right) + \pi a^2 \rho g \sin^2 \phi (a \cos \phi - y)$$
$$- 2\pi a T \sin \phi \cdot \sin (\theta + \phi),$$

where ρ is the density and T is the surface tension of the liquid.

If the sphere is immersed completely, the upthrust becomes

$$U_1 = \tfrac{4}{3} \pi a^3 \rho g.$$

Hence, in general,

$$\frac{U}{U_1} = \left(\frac{1}{2} - \frac{3}{4} \cos \phi + \frac{1}{4} \cos^3 \phi \right) \mp \frac{3}{4} \sin^2 \phi \left(\cos \phi - \frac{y}{a} \right)$$
$$- \frac{3}{2} \cdot \frac{T}{a^2 \rho g} \cdot \sin \phi \cdot \sin (\theta + \phi). \quad \ldots \ldots (1)$$
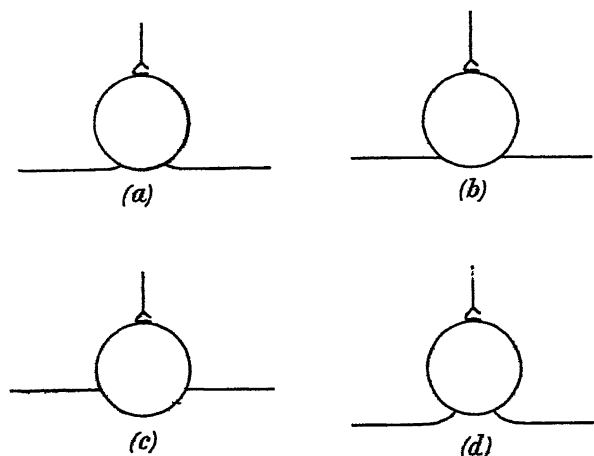
Figure 2.

(a), (b). Mercury rising.   (c), (d). Mercury falling.

In this expression for U/U_1, the first term is always positive whatever the value of ϕ. The second term is positive or negative according as $\cos \phi$ is greater or less than y/a, i.e. according as $(\theta - \phi)$ is greater or less than π. Clearly, also, the same condition determines whether the effect of the third term is positive or negative.

Let it be supposed that the sphere is immersed so that the mercury is exactly horizontal right up to the lines of contact (figure 2 (b) or 2 (c)). It follows then that $(\theta + \phi) = \pi$ and that $\cos \phi = y/a$. If the upward force acting on the sphere under these conditions is denoted by U_0, then

$$\frac{U_0}{U_1} = \frac{1}{2} - \frac{3}{4} \left(\frac{y}{a} \right) + \frac{1}{4} \left(\frac{y}{a} \right)^3 \quad \ldots \ldots (2)$$

This criterion makes it possible to recognize the point at which the liquid surface is horizontal right up to the surface of the sphere much more accurately than by direct visual observation. We have, therefore, a dynamical method of observing angles of contact which is entirely independent of the surface tension of the liquid.

§2. EXPERIMENTAL METHOD

Calculation suggests that the most suitable diameter of sphere is about 1 cm. In some of the experiments, steel ball-bearings were used ; in others, small glass spheres suitably weighted with mercury before being sealed. Each sphere in turn was suspended from a Sucksmith ring balance (Sucksmith, 1929) of simplified design, by the displacement of which the force, upwards or downwards, exerted by the mercury could be measured. The ring was of steel tape approximately 0·008 cm. in thickness, 0·4 cm. in width and 5 cm. in diameter. Using a single mirror in conjunction with a telescope and a scale at a distance of 2 metres, a force of 1 gm. wt. gave a change in reading of approximately 4 cm., while the scale could be read to one-tenth of a millimetre.

All grease was removed from the steel spheres by washing three times in ether (B.P.). In some of the experiments the glass spheres were prepared in the same way. In other cases they were cleaned in concentrated nitric acid, washed in many changes of distilled water, and finally allowed to dry under cover to prevent contamination by dust. Redistilled mercury, filtered to give a clean new surface at the beginning of each set of observations, was contained in a square glass trough of side 7 cm., sufficiently large, that is, to give a truly plane surface. By means of three similar calibrated screws, the mercury surface was brought up into contact with the lowest point of the sphere, then slowly raised in steps of about 0·01 cm. so as to immerse about half of the sphere, and subsequently lowered until contact was again broken. The time taken for a total motion of a few millimetres was rather more than an hour. The height y of the centre of the sphere above the plane surface of the mercury throughout this procedure was determined from the rotation of the screws and the small change in the absolute position of the sphere as the ring balance was deflected.

The variation of the quantity U/U_1 with the ratio y/a during the raising and lowering of the mercury is shown by the full curves in figure 3 for the case of a steel sphere of radius 0·635 cm. On the same diagram the thin curve represents the variation, with the ratio y/a, of the quantity U_0/U_1, as calculated from equation (2). The points of intersection of the experimental and theoretical curves give the values of y/a when the mercury surface is horizontal right up to the surface of the sphere. The angles of contact are then given by the equation

$$-\cos\theta = \cos\phi = y/a.$$

Clearly the method is capable of considerable precision.

It will be observed that, on first making contact ($y/a = 1$), the mercury rises a short distance up the curved surface, as shown in figure 2 (*a*), and so exerts a downward force on the sphere, the second and third terms in equation (1) being negative and together greater than the first. This adhesion of the mercury to the solid is only observed with really clean surfaces. Adhesion takes place similarly when the mercury is falling, but now, the angle of contact being smaller, quite a pronounced neck develops (figure 2 (*d*)). The downward pull on the sphere passes through a maximum value, as in Ferguson's experiments (1913), after which contact is broken, and a minute drop of mercury remains at the lowest point of the sphere, which may now be as much as a millimetre above the level surface of the mercury.

§3. RESULTS

The table shows the measured angles of contact for a number of different spheres. Although in any given case the angles of contact may be measured quite accurately, the angles themselves are not exactly reproducible, even when scrupulous cleanliness is observed. It will be noticed, however, that the receding
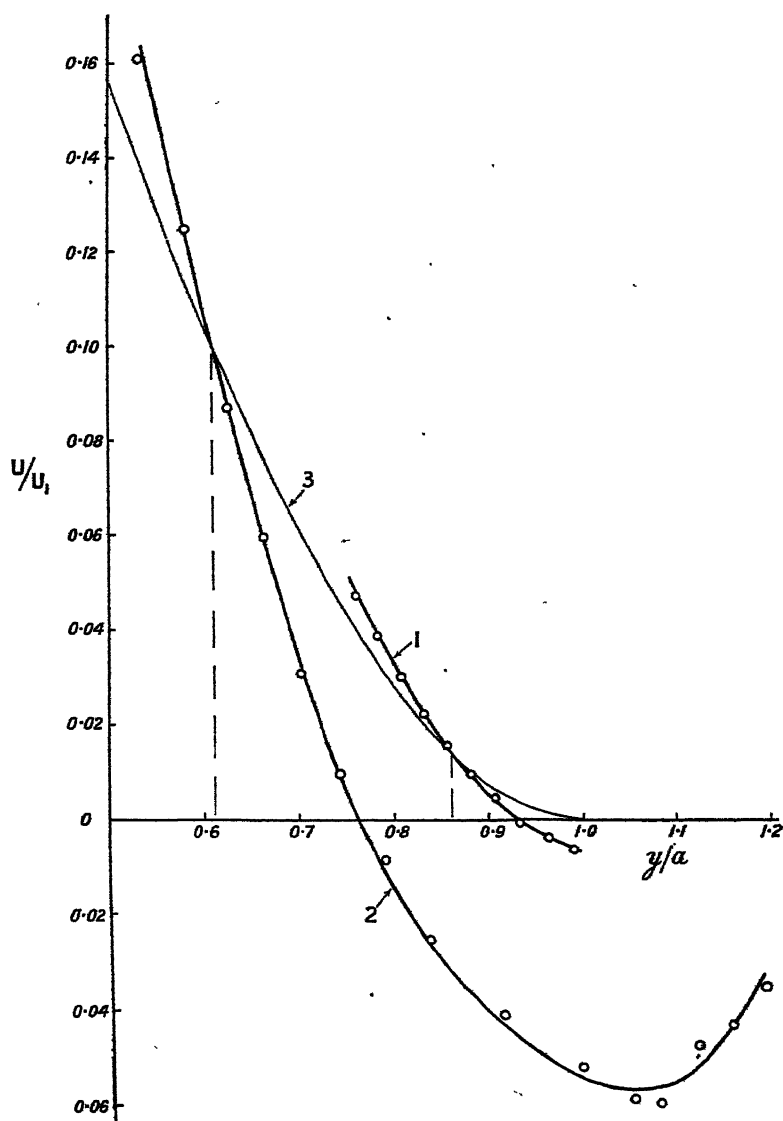
Figure 3. Force on steel sphere, $a=0.635$ cm.
1. Mercury rising. 2. Mercury falling. 3. U_0/U_1 (calculated).

angles all lie in a comparatively narrow range, while the advancing angles vary much more widely. The difference between the two angles is greater when the glass spheres are washed in ether than when they are washed in nitric acid and distilled water. No trace of residue, however, was visible on evaporating several

cubic centimetres of ether in a watch-glass. In no case, however, is the difference between the cosines of the angles of contact as great as the difference observed for a mercury index travelling in a glass tube at a velocity of 3 cm. per sec. or faster (Yarnold, 1938). The first group of steel spheres gives angles of contact hardly different from those observed with glass spheres cleaned in the same way. It must not be inferred, however, that this result is due to the presence of a film left by the ether, which forms the true surface in each case, since two other steel spheres, obtained from a different source, give much higher advancing angles.

Following Adam and Jessop (1925), it is usual to ascribe the hysteresis of the angle of contact of a liquid to a frictional force F per unit length acting at right angles to the line of contact. The force F is tacitly assumed to be the same whether the liquid is advancing or receding over the surface of the solid, and, as a consequence, the cosine of the true equilibrium angle of contact is taken to be the mean of the cosines of the advancing and receding angles. The introduction of the force F, however, is equivalent to the statement that part of the

Material	Radius (cm.)	Remarks	$-\cos\theta$		θ (degrees)	
			Advancing	Receding	Advancing	Receding
Glass	0·574	Washed in nitric acid	0·723	0·600	136·5	127·0
,,	0·522	and distilled water	0·780	0·596	141·0	126·5
Glass	0·574	Stored in alcohol and	0·831	0·648	146·0	130·5
,,	0·522	washed in ether	0·848	0·600	148·0	127·0
Steel	0·635	Washed in ether	0·860	0·610	149·5	127·5
,,	0·357		0·806	0·575	144·0	125·0
,,	0·318		0·857	0·648	149·0	130·5
Steel	0·556	Washed in ether	0·930	0·624	158·5	128·5
,,	0·437		0·967	0·640	165·0	130·0

work expended in the formation of fresh surface is irreversible. The assumption that the magnitude of F is independent of the direction of motion of the line of contact is therefore equivalent to the assumption that the irreversible work expended in the formation of unit area of the solid-air interface is the same as the irreversible work expended in the formation of unit area of the solid-liquid interface.

Stated in this form, the assumption hardly appears to be justified, though the error may not be serious when the two angles of contact differ only slightly. Nor does it appear to be supported by the measurements of the angles of contact of mercury. The approximate constancy of the receding angle, together with the remarkable adhesion of the mercury to the solid surface, suggests that a solid-liquid interface of more or less constant energy is formed in all cases, and that more or less the same quantity of irreversible work is expended per unit area in its destruction. On the other hand, the wide variation in the advancing angle

suggests that the irreversible work involved in the destruction of the solid-air interface and its replacement by the solid-liquid interface may be extremely variable.

REFERENCES

ADAM, N. K. and JESSOP, G., 1925. *J. Chem. Soc.* **127**, 1863.
FERGUSON, A., 1913. *Phil. Mag.* **26**, 925 ; 1914. *Ibid.* **28**, 149.
SUCKSMITH, W., 1929. *Phil. Mag.* **8**, 158.
YARNOLD, G. D., 1938. *Proc. Phys. Soc.* **50**, 540 ; 1940. *Ibid.* **52**, 191.

ADDRESS

DELIVERED AT THE OPENING OF THE THIRTIETH EXHIBITION OF SCIENTIFIC INSTRUMENTS AND APPARATUS, 1 JANUARY 1946, BY

THE RIGHT HON. SIR STAFFORD CRIPPS, P.C., M.P.

THE Physical Society is much to be congratulated upon its enterprise in organizing an exhibition of scientific instruments, thus giving an opportunity to British scientific instrument makers to demonstrate once again the excellence and ingenuity of their workmanship.

During the years of war we have of necessity become more self-reliant for our research and development in all branches of science and technology upon our own native manufactures. And judging by the outstanding results which we have achieved during the last six years there can be no doubt at all that our scientific instrument makers have served us well. Many of them have turned their efforts to the direct production of instruments for aircraft, tanks or ships of all kinds and have thus entered the field of semi-mass production or even of mass production itself, but this has not prevented them from continuing with the development and production of those many scientific instruments of extreme accuracy upon which so much of our present research and industrial technique is based.

Indeed, instead of debasing their standards by introducing volume production they have carried into that mass production the skill and accuracy which they had long cultivated in their specialized and small quantity productions.

In the result the war should have put us into a position to expand our scientific instrument production upon a large scale, preserving all the excellencies for which our leading firms have so justly been famed in the past.

And the present moment is a most appropriate one in which to demonstrate this fact not merely to our own users of these instruments but also to the potential customers for them from all parts of the world.

As in the first world war, so in this last one, we have learnt how backward we have been in our research efforts during times of peace. Research is easily and widely recognized as the sole basis for rapid progress and superiority in wartime; so much so that every effort is made by the State to entrain every scientist available in one line of research or investigation or in another. The scientist is the blue-eyed boy called in to help the solution of every problem and to force the pace of progress in a never-ending duel of wits and intelligence with the enemy.

And the scientist must be duly and properly equipped for his work.   Great research establishments are improvised regardless of cost and filled with the most up-to-date instruments and machinery—which, incidentally, put to shame the accommodation and facilities of our peacetime universities and other research centres.

Nothing is regarded as wasted that may contribute to the superiority over the enemy in any field.   Tens of thousands of research man-hours are gladly sacrificed to the merest chance of some new development which will give us advantage over the foe.

But as we experienced between the two wars, the atmosphere is apt to change when the crucial danger of war is past.   We then weigh up the £ s. d. and evaluate whether this or that expenditure is worth our while.   And in that calculation we tend to become shorter and shorter sighted.   We are liable to forget that progress and superiority in peace are as much based upon research as they are in war. The tempo of our drive forward on the scientific front slows down and we fail to educate and train a sufficiency of highly skilled staff to maintain the intensity of our research.

That at least was the experience after the last war, and it is a folly that we must at all cost avoid repeating after this war.

Research is the life-blood of our industrial progress, and industrial progress is the basis for our continuing prosperity as a great exporting nation.

To attain the degree of research activity which is essential to our future as a great inventive industrial nation four things are required.

First we must have the men and women adequately trained and in sufficient numbers.   That is a matter of more and better educational facilities of all kinds and a free access to those facilities for the intelligent youth wherever they may be and whatever kind of homes they come from.

Second we must raise the status of our scientists.   Too long they have been regarded as rather superior craftsmen definitely inferior to the leaders of industry or the principal administrators.   The Classical Tradition in education still claims a superiority which is out of date, and the scientist is still looked upon by many as of a lower grade.   Whether we look at the salaries they are offered or the position to which they are relegated in the hierarchy of industry or of the State Service, we see that they are kept down in a position of inferiority.

I am not one of those who claim that the scientists should rule in all spheres, nor am I a believer in technocracy, but I am firmly convinced that both scientists and technicians are worthy, and have amply proved themselves worthy, of a position equal to that of any other section of the community.

But until we gain the recognition of this fact by the country as a whole we shall not get the best out of our scientists, nor shall we make the headway that we should in applying the results of our research to our every-day needs.

The scientist must not be regarded as an individual to be called in to assist in an emergency when matters have got into a hopeless position—like the plumber with a burst water-pipe—rather he should be looked upon as a valuable and equal co-operator in every phase of the solution of our problems of living.

Third, we require proper accommodation and facilities for our research.   It is not necessary that every research institution should be a palace with chromium-

plated ware, nor do I believe that string and sealing wax provide the essential setting for all that is of value in research. There is a happy mean in which men and women can do their best work. provided with the most up-to-date instruments that they require.

A good many of our research establishments today fall far below that mean in standard both as regards space—and adequate space is vital to good research—and as regards equipment.

I hope that full advantage will be taken of wartime surpluses to see that all our universities and other research establishments are fully equipped with adequate machinery and instruments.

Finally, there is the all-important matter with which this exhibition is concerned. We cannot afford to be dependent upon others for the instruments necessary for our education of scientists or for the research work which they do, nor, indeed, for all those many testing instruments which are today part and parcel of every up-to-date industrial concern. It is essential that we should have and maintain in this country a pre-eminent scientific instrument manufacture both for our own use and to help to supply those many other countries who cannot themselves afford to set up such manufactures.

I believe that if we could insist upon these four points we could lay the permanent and solid foundation of a live, active and vigorous research in our country which would enable us to keep in the van of the great industrial countries of the world.

It is because I believe that this exhibition will help along the lines that I have sketched that I am so delighted to have the honour today of declaring this exhibition open.

REVIEWS OF BOOKS

Elementary Wave Mechanics, by W. Heitler. Pp. viii + 136. (Oxford and London : Sir Humphrey Milford, at the Oxford University Press, 1945.) 7s. 6d.

In this little book, Professor Heitler gives the substance of a course of lectures which he delivered in Dublin. It is truly elementary, both in the demands which it makes on the previous knowledge and mathematical ability of the reader, and also in that it deals only with the *elements* of wave mechanics. There is here no metaphysics, but a plain introduction to the physical arguments which led de Broglie and Schrödinger to the wave equation. The uncertainty principle is introduced as a demonstrable theorem, and not as a vaguely supported principle, and is indeed at first named the uncertainty relation, a change in nomenclature which might be advantageous if generally adopted.

The wave equation is used to investigate the hydrogen atom, the Zeeman effect and the two-electron problem in a manner which is not now unfamiliar in elementary texts. The Pauli exclusion principle is introduced in the last-named chapter.

The best part of the book, however, is in the last three of the nine chapters, where the periodic system, the theory of the homopolar chemical bond and the general problem of valency are taken up. In conformity with the plan of the book, these matters are treated in part in qualitative fashion, with guidance from the theory of simple cases, but the treatment is so lucid that these chapters must be of value to all who are concerned with the problems of chemical combination. The explanation of how it comes about that valencies can become saturated after a limited number of atoms have combined, or of how the carbon atom lends itself to the multitudinous compounds in which it is

found, are of more value for a clear picture than would be many pages of detailed algebra working out a more complete theory.

In reviewing the same author's *Quantum Theory of Radiation*, the view was expressed that the author would do a service if he wrote a more elementary introduction to the subject. Can he now put all students in his debt by writing a book intermediate in standard and difficulty between this very elementary treatment of the simplest wave mechanics and that very difficult work in which the whole range of quantum mechanics is called on ? Such an intermediate treatment would, presumably, deal fairly exhaustively with the mechanics of particles and would have some introduction to the subject (untouched in the elementary book) of the interaction between matter and radiation, but without attempting an exhaustive treatment of this more difficult topic.   J. H. A.

Questions Actuelles de Géophysique Théoretique et Appliquée, par EDMOND ROTHÉ. (Paris : Gauthier-Villars, 1943.) 275 Fr.

Written and published during the German occupation of France, this book has only recently reached Great Britain. It covers a very broad field, and from the data that have been obtained in the realms of gravity observations, radioactivity of the earth's crust, volcanology, seismology and terrestrial magnetism an effort is made to present a unified picture of the modern conception of the earth's constitution. Isostatic compensation is discussed in the light of gravity observations, and is supported by some detail of the work of the Swedish geologist de Geer on varved clays and the retreat of the ice-sheet, leading to the subsequent elevation of the land masses due to the removal of the ice-load. The two following chapters on radioactivity and volcanic action are broadly discussed in relation to isostasy, so that these branches are integrated into a self-consistent whole.

In view of the unity up to this point, it is rather surprising to find the section on seismology so short, and in particular the importance of near earthquakes, and the information revealed by them on the sequence and thickness of layers in the crust, so closely allied to the previous aspects, is scarcely mentioned. The chapter on terrestrial magnetism is also disappointing and does not appear to fit the main theme established in the early part. After a brief description of the earth's field, and ending with the diurnal and annual variations, the main part is devoted to a speculative theory, due to Haalek, on the cause of the earth's magnetism. The work on isopors and its possible implications is ignored. At the end of each chapter a short section describes some of the practical applications to geophysical prospecting, and here the contents supplement a previous publication, devoted to these aspects, by the same author.

A number of numerical errors occur in the text ; but in the majority of cases they are obvious, and will, no doubt, be corrected in later editions. Some of the illustrations leave much to be desired, the reproduction is poor and a few are not easy to follow. In figure 67, for example, the axes are not labelled, but the fortunate inclusion of a table, giving the same data, allows them to be disentangled. In spite of these defects the book is a good review of the present ideas on the subjects treated.   J. M. B.

The Mathematics of Ultra-high Frequencies in Radio, by L. N. BRILLOUIN. Pp. vi + 210. (Brown University, Providence, R.I., 1943.) Price not stated.

The theoretical background to micro-wave radio has become of increasing importance during recent years, and the present book, which is a record of a course of lectures given at Brown University in 1943, is a welcome addition to the few existing works dealing with the theory. It provides a unified account both of the physical concepts associated with wave propagation and the mathematical methods used in analysing the electromagnetic field. It is readable by the physicist and engineer as well as the mathematician ; and by seriously studying the book, the student and the research worker will avoid the necessity of consulting a large number of the original papers which lie scattered throughout the scientific literature of the past decade.

The book begins at an elementary level. Tuning circuits with lumped or with distributed constants are discussed, and the usual introduction to phase and group velocity

is given. Transmission lines and co-axial cables are studied and there is a section on the various types and relationships of units. Acoustic waves are studied preparatory to electromagnetic waves in pipes.

Then the mathematics proper is introduced. The author assumes slight initial knowledge on the part of the reader and begins with a discussion of vector algebra, scalar and vector fields and the expression in curvilinear co-ordinates of various vector functions. Maxwell's equations are set up and solved by the method of the retarded potential. The expression of these equations in tensor form (fundamental from the relativistic point of view) is given, and the Hertz potential and Poynting vector are discussed. A short chapter is devoted to the skin effect.

The propagation of waves along a cylinder of arbitrary cross-section is studied mathematically in some detail ; the treatment is general at first, but particular cases discussed include tubes of rectangular, circular or elliptical cross-section and co-axial cables. All these cases lend themselves readily to a mathematical analysis, and the various types of possible waves are classified.

The last chapter is given to cavity resonators. It is a pity that the treatment is not more comprehensive, and while attenuation, damping and the distribution of the higher modes are discussed, no mention is made of some of the practical problems in resonator theory, such as the calculation of resonant frequencies and losses in a compound resonator by the (numerical) relaxation method, or, in special cases, by the Hahn method of matching solutions. Similarly, in the chapter on wave guides no mention is made of discontinuities in the cross-section, although this has been the subject of discussion in several papers appearing in recent literature. It is disappointing to find no mention of antenna theory (except the reciprocity theorem). But here again recent work by the author on this subject may be found in the literature.

The lecture notes are well presented and the student should not find it difficult to read them. It is gratifying to note the enterprise of Brown University in making available, in mimeographed form, the results of the Advanced Instruction and Research in Mathematics. L. S. GODDARD.

Science in Progress. Fourth Series. Pp. 331. (Yale : University Press ; London : Sir Humphrey Milford, at the Oxford University Press, 1945.) $3.00.

This delightful set of essays is made up of contributions from eleven very distinguished specialists in various branches of knowledge, edited by George A. Baitsell and with a foreword by L. L. Woodruff. In his thoughtful foreword Mr. Woodruff says :—" The accumulation of knowledge during the millennia since man's emergence from a condition of complete dependence on his environment to the relatively masterful position that he now holds is, except in its broad sweep, beyond the compass of one mind, however talented. The polyhistor has been succeeded by the specialist." This is indeed true, and for this reason the reviewer feels quite unable to do much more than give a broad picture in bold outline of the various subjects presented.

The contributions are mainly the substance of national lectures delivered under the auspices of the organization known as " The Society of the Sigma XI ", which is devoted to the encouragement of research in science, and the book represents the fourth series of such lectures.

The collection is intended to provide scientific men with general outlines, although accurate ones, of fields of research in which, while not actually working, they may be interested. From this point of view the purpose in view has been excellently achieved.

Each author has provided a most excellent series of references attached to his chapter, and it is flattering to national vanity to see that so many Englishmen are mentioned as having pioneered or followed the exciting hunts for the new knowledge which goes to the making of the various chapters.

A striking war-time chapter is the first, by Professor Miles of Yale. The importance of the psychological approach to certain problems in military aviation has been proved

to be of paramount importance, and the interested reader will learn how an aviation cadet is examined from all points of view so as to ensure not only that he possesses the personal qualities so essential for an airman's success, but also that he receives the training to develop fully these qualities.

In the second chapter Professor Bronk of the University of Pennsylvania speaks also of war-time interests in the combined fields of Physics and Biology. He gives attention to the subject of sensory impulses from a pressure-sensitive nerve ending, and there is much interesting matter belonging to past history and the live present in regard to those rhythmic oscillations of potential in nerve fibres which are so important in impulse propagation along nerves.

Professor Hecht of Harvard follows with an extraordinarily interesting account of the amount of energy required in the process of vision. He explains the experiments which have led to the remarkable result that for a normal individual to perceive a faint flash of light on the threshold of visibility the minimum energy lies between 5 and 8 quanta.

Professor Loewi of New York University, formerly of Gratz, gives a fascinating account of the discovery of the chemical transmission of nerve impulses. It is, indeed, remarkable that the nervous system produces so many of its effects through the liberation of chemical substances.

From the pen of the late Professor Birkhoff comes a masterly historical account of some aspects of mathematical physics garnished with personal recollection and luminous with individuality—it makes excellent reading.

Professor Debye of Cornell gives, as is to be expected, a first-rate résumé of modern methods of obtaining very low temperatures and some of the problems to which these methods have provided a key. The subject of his choice is clearly not a closed book, for he points out the possibility of extending the experimentally accessible range towards the absolute zero by removing the disorder which may exist in the nucleus itself and connected with nuclear magnetism, just as the methods he described have nearly removed all the disorder associated with atoms and their motion.

Professor Eyring of Princeton contributes a valuable article on physical chemistry, including a consideration of the distribution of energy in a molecular system, rates of reaction, and luminescence inhibition.

Professor Rabi of Columbia discourses on molecular beams and " radio-frequency spectroscopy ".

Dr. Hickman, who is Director of Research of the Distillation Products Company, and is an Englishman, produces a valuable chapter on vacuum chemistry—in particular molecular stills and their applications, for example in the vitamin A industry. The new substance kitol, which is the mirror image in properties of β-carotin, itself the normal precursor of vitamin A, was discovered with the use of the vacuum still.

Professor Elvehjem of Wisconsin gives a detailed and convincing account of the vitamin B complex. The formulae of the various substances, about a dozen, are discussed where possible and the dietary factors explained—altogether a most illuminating article.

Professor Cohn of Harvard University, in this last chapter, gives a most interesting account of work on which he and his colleagues have been engaged during the war on the chemistry of proteins, particularly in the chemical interactions and the separation of the protein components in protein solutions—researches which involved the proteins of blood plasma, so important in the alleviation of human suffering. A great deal of information appears to be concentrated in a few pages, from the relative dimensions of various proteins through the chemical fractionation of the components of plasma to details in connection with the regulation of bodily function by hormones and enzymes.

The impression gathered from reading through this remarkable collection of masterly essays is one of immense satisfaction at being privileged in one small volume to read such clearly presented accounts of current research in so many widely separated fields. The references at the end of the book will be valuable to anyone whose curiosity is aroused sufficiently to follow up particular points ; the excellent diagrams and photographs are fully illustrative and give the impression of great accuracy. If one may intrude a word of friendly envy, it is to say what a rare pleasure it is to handle a book produced so lavishly and so perfectly on such beautiful paper. If it can be bought in this country, it is magnificent value for three dollars.   R. WHIDDINGTON.

Chemical Crystallography, an introduction to optical and X-ray methods, by C. W. BUNN. Pp. xii+422 and 13 plates. (Oxford : Clarendon Press, 1945.) 25s.

The book is addressed to " students of chemistry who wish to acquire some knowledge of crystallographic methods and research workers who wish to make practical use of such methods ". It is divided into two sections, the first dealing with the identification of solid substances by physical methods and the second with the determination of crystal structures. After an introductory survey, the author begins with a chapter entitled *The Shapes of Crystals*, which gives an introduction to crystal morphology and the relation between shape and structure. He approaches the subject of crystal symmetry through the phenomena of crystal growth and achieves a simplified treatment of elementary crystallography, e.g. indices of crystal planes by means of the unit cell rather than in terms of axial ratios. This is followed by an account of crystal optics and the properties of crystals observed in the polarizing microscope, leading up to the procedure which is then given for the identification of transparent crystals by microscopic methods. The last chapter of section one introduces the elementary theory of diffraction of x rays and is concerned with the practical details of taking x-ray powder photographs for identification.

Over half the book is devoted mainly to the determination of the arrangements of atoms in crystals. This second section opens with an account of the single-crystal techniques, how unit cell dimensions are measured and how this knowledge is applied, amongst other things, to identification and determination of molecular weight. Succeeding chapters give a masterly treatment of the principles involved in the determination of crystal structures by trial and error, and numerous examples of both organic and inorganic compounds are worked out in detail. A chapter on electron density maps and vector maps includes an account of Fourier series methods, of the use of optical methods in place of calculations and the scope of vector methods. The author also discusses the evidence on crystal structure given by optical, magnetic and other physical properties, and concludes with a brief account of broadened x-ray reflections and diffuse spots.

The well-balanced review of progress in method and technique since the publication of *The Crystalline State*, by W. L. Bragg, makes this an indispensable companion to that volume. Formal physical and mathematical treatment are avoided, but the book has a discipline of its own and demands careful reading. The author assumes more than a passing interest in both the organic and inorganic compounds, thus revealing the range of his own contributions to the subject. He has not hesitated to adopt an unorthodox development in order to reflect the course of his own experience in applying crystallographic methods to chemical problems.

The text is profusely illustrated and provided with useful appendices and excellent indexes. Both author and publisher are to be congratulated on producing so fine a book under war conditions at a reasonable price. F. A. B.

Polarographic and Spectrographic Analysis of High Purity Zinc and Zinc Alloys for Die Casting, by various authors. Pp. 117. (H.M. Stationery Office, 1945.) 5s.

An account is presented of researches, carried out under a British Standard Institution Panel, to investigate the use of physico-chemical methods for the analysis of high-purity zinc and zinc alloys for die-casting. The work concerned has proved the basis for a British Standard Method on the same subject, and the present publication is in the nature of a detailed discussion of alternative techniques, rather than a working manual for use in the laboratory. Such being the case, its primary appeal is to the specialist and to the spectrographer in particular, but the body of evidence presented is sufficient to warrant consideration by all interested in analysis in the wider sense. In view of the emphasis in modern spectrographic research on the investigation of different excitation methods, two papers giving considerable detail of variations on simple condensed sparks and an intermittent A.C. arc technique are of considerable interest. In a further section an arc technique is described in which the sample is placed on a suitable electrode in the form of oxide: This procedure allows of the

preparation of standards of known composition, enabling the spectrographer to be independent of chemical analysis. In this paper methods of plate calibration are discussed in some detail, but at the same time it is agreeable to find the results presented in such a manner as to be readily intelligible to a non-specialist reader.

The description of an alternative method of attack by the polarograph will be of interest to laboratories without spectrographic equipment, and also as providing an independent result by an entirely different technique. It is shown that results of considerable precision can be achieved by modifications in chemical pre-treatment, suitable to the particular problem in hand.

In fact, the standard deviations of the results by all the methods discussed agree very closely and provide important additional evidence, not in vague generalizations but in a precise and concrete manner, of the value of physical methods to the critical analyst. Finally, it may not be out of place to mention that the amount of information presented for a modest sum compares exceedingly favourably with the usual scientific text-book.

H. G. SHORT.

Radiocristallographie, by ANDRÉ GUINIER. Pp. ix + 294, with 15 plates. (Paris: Dunod, 1945.) Price not stated.

This can be recommended as a really excellent book, the sort of book that has been badly needed for several years. It is written specifically for those who wish to use the study of crystals by x-ray methods as a means to some further end. It is not meant for specialists in crystal-structure determination. Yet all will benefit by reading it, especially those who lightly play with structure factor tables, Fourier series and calculating machines, without ever troubling to understand the foundations upon which the subject is built. True, it dismisses gas tubes in a contemptuous seven lines and gives 1 in 10,000 as a desirable limit for precision work with powder cameras. But it introduces the reciprocal lattice where it ought to be introduced—right at the beginning; it relegates all tiresome but necessary mathematical proofs to appendices; it show the *raison d'être* of some of those mysterious tables in the *International Tables for Crystal Structure Determination* that we ourselves never seem to need; and it lays proper emphasis on the study of crystal texture and imperfection. Every theorem is illustrated by practical and interesting examples; and numerous useful hints about apparatus and methods are given from the author's own wide experience.

Yet the book gives a bad first impression. The pages are uncut and the title is not clear to an English technical reader. It sounds as if it might have something to do with wireless receiving sets, but the prospective buyer is left to guess. After all, x rays are not entitled to a corner in the expression "radio". The cover is of thin paper of a bilious shade of yellow (or is the reviewer perhaps jaundiced through having read the book during an attack of gastric influenza?). The paper is of extremely poor quality and bad colour; the type is crowded and difficult to read. There is a table of contents, but no subject or author index, and the references are not up to date (for instance, the reader is referred to the 1932, instead of the 1940, edition of Clark's *Applied X-Rays*). Finally, and most unkindest cut of all, the present writer's name is misspelt. It was Robert Louis Stevenson who, when some American publisher (*before* the days of lease-lend) had pirated one of his books, said that it was not the theft that angered him so much as the mis-spelling of his name. "I saw my book advertised as the work of R. L. Stephenson", he says, "and I own I boiled. It is so easy to know the name of a man whose book you have stolen, for there it is, full length, on the title-page of your booty. But, no, damn him, not he! He calls me Stephenson".

But in this case the author and publisher must be forgiven. The book is a revelation, not only of the difficulties of writing and producing in occupied France, but of the indomitable spirit that would undertake such a work at such a time and under such conditions. And, with all its faults, it is certainly a book that will be referred to again and again. The pity is that it will fall to pieces in the process; but surely we may soon expect a more worthily produced second edition, with references brought up to date and a modernization of the text where necessary? Or cannot some British or American publisher negotiate for the translation rights and produce an agreed English text in a [illegible] it deserves it.

K. LONSDALE.

CRYSTAL STRUCTURE OF DOUBLE OXIDES OF THE PEROVSKITE TYPE

BY HELEN D. MEGAW,
Birkbeck College

ABSTRACT. The cell dimensions of a number of double oxides belonging to the perovskite type have been accurately determined from examination of high-angle lines on x-ray powder photographs. The structures found fall into groups, as follows :

(1) *Cubic (ideal perovskite type).* This includes $SrTiO_3$, $SrSnO_3$, $SrZrO_3$, $BaSnO_3$, $BaZrO_3$, $BaThO_3$; also $BaTiO_3$ above 120° c.

(2) *Tetragonal.* This includes the usual form of $BaTiO_3$ at room temperature, $PbTiO_3$, and $PbZrO_3$. The unit cell is derived from the cubic by simple extension or compression along one tetrad axis ; and, like the cubic, it contains the formula number of atoms.

(3) *Orthorhombic.* This includes $CaTiO_3$ (the mineral perovskite), $CaSnO_3$, $CaZrO_3$, and $CdTiO_3$ (fired above 1100° c.). The structure is derived from the cubic by a shear in the (010) plane and a slight extension or compression along the b-axis, giving a monoclinic pseudo-cell with *a* and *c* edges equal. The lattice is thus actually orthorhombic, and should be referred to new *a* and *c* axes lying at approximately 45° to the old in the same plane. There is an observed doubling of the cell edges, attributable to changes in some of the atomic parameters.

(4) *Rhombohedral.* $BaTiO_3$ can be prepared in this form, though the conditions are not yet fully established. The pseudo-cell is obtained by a very slight compression of the cubic cell along the cube diagonal ; the true cell is a multiple of this.

Steric considerations, based on Goldschmidt's ionic radii, are used to account for the occurrence of the different structure modifications.

§ 1. INTRODUCTION

THE mineral perovskite, $CaTiO_3$, has given its name to a very simple and important structure type, classified as E 21 in the *Strukturbericht*. This is cubic, with one formula–weight per unit cell. A large number of double oxides (as well as some iodates and double halides) are classified as belonging to this type. Early work on these, by Goldschmidt (1927) and his school, was directed to illustrating the two principles which are characteristic of multiple isodesmic structures, namely, the dependence of structure type on ionic radii and its relative independence of valency.

There was no emphasis in Goldschmidt's work on the detailed geometry of the individual structures. It was recognized that while some, such as $SrTiO_3$, were strictly cubic, others, such as perovskite itself, must from the crystallographic evidence be of lower symmetry, though the difference was not detectable by x-ray methods; while others again, such as $CaSnO_3$, showed departures from

cubic symmetry in the geometry of the unit cell. Advances in x-ray technique in the intervening years now make possible more accurate determination of the cell dimensions and a more critical discrimination between the cubic and various " deformed " structures. Some work has already been done along these lines. Hoffman (1935) investigated a number of compounds, including some which had not been previously prepared, and classified them into " cubic " and " deformed " without attempting to describe the deformation in detail. Naray-Szabo (1943 a) made a complete investigation of the structure of perovskite, both single crystals and powder, and found the structure monoclinic; his results will be discussed in more detail later. He investigated a number of other related compounds (1943 b), distinguishing those which belonged to the same " new type " as perovskite from those which were strictly cubic, and showing their relationship to cubic ReO_3 and rhombohedral ScF_3. Recently the present author showed that $BaTiO_3$ possessed a tetragonal modification of the structure (Megaw, 1945). This was also found by Rooksby (1945), who gave a brief discussion of some of the other non-cubic structures.

It is clear that this group of compounds includes structures with a very varied symmetry, all based on small but completely different modifications of the same cubic elementary cell.

A comparison of the different variants, and an attempt to understand why one rather than another should occur in a given case, is likely to be useful in any further work on the nature of forces in ionic solids. The first step towards this is an accurate determination of the geometry of the unit cells of the different structures, which was the object of the present investigation.

§2. DESCRIPTION OF STRUCTURE

It seems convenient to define the term *perovskite type* to include all structures which retain a pseudo-cell derived from that of the ideal cubic structure E 21 by small distortions either of cell edge or of interaxial angle. The true cell may require a doubling of one or more edges of the psudo-cell, or a different choice of axes. When necessary, the cubic structure E 21 can be distinguished as the *ideal perovskite type*. Rooksby suggests that these structures should be called *pseudo-isomorphous*, but here the word *homotypic* is preferred, as it lays emphasis on the common crystal type, and does not carry the chemical implications involved in the use of the older word.

Compounds belonging to the perovskite type have in general the formula ABO_3, where the sum of the cation valencies is 6. (ReO_3 and WO_3 may be included as extreme cases in which one set of cation sites is completely unoccupied, and the sodium tungsten bronzes, Na_xWO_3, are an important example of a defect lattice with variable tungsten valency. Naray-Szabo classes all these as " sister-structures "). The present work deals only with the numerous group in which A is 2-valent, B 4-valent.

In the ideal type, the atomic parameters are as follows:

$$A \qquad \tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2}$$
$$B \qquad 000$$
$$O \qquad \tfrac{1}{2}00, \ 0\tfrac{1}{2}0, \ 00\tfrac{1}{2}$$

This gives a structure in which the B cations are surrounded by regular octahedra of oxygens, which are linked together by their corners to form a three-dimensional framework. The A cations occupy the large holes between the octahedra, where each is surrounded by 12 oxygens. The A-O and B-O electrostatic bond strengths (in Pauling's sense) are thus $\frac{2}{12}$ and $\frac{4}{6}$ or $\frac{1}{6}$ and $\frac{2}{3}$ respectively. There are no bond strengths equal to or greater than 1, and hence the structure may be classified as isodesmic (Evans, 1939). Chemically this implies that the compounds should be described as double oxides rather than as titanates, zirconates, etc., but the latter nomenclature is sometimes convenient and not seriously liable to mislead. Figure 1 (taken from Nara-Szabo's paper) shows the structure.

A^{2+} ion

B^{4+} ion

Oxygens at corners of octahedra.
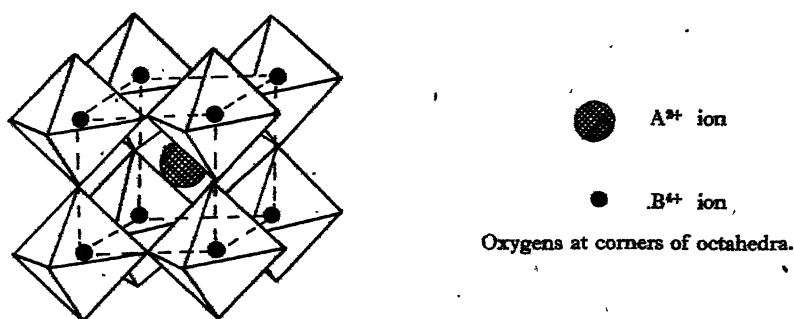
Figure 1. Ideal cubic perovskite structure.

In all the structures investigated, this topographical description remains true, though in many of them the framework loses its symmetrical construction. Two main types of deformation are possible, and may occur separately or together.

(1) The unit cell changes its shape, by altering either the relative lengths of the cell-edges or the axial angle.

(2) The atomic parameters of some or all of the atoms are slightly altered. Since in the ideal structure the atoms are in special positions, slight displacement in any direction involves a lowering of the symmetry, and in general also a doubling of at least one cell edge. Conversely, a doubled cell edge implies atomic parameters different from those of the ideal structure.

In the present work, no systematic attempt has been made to determine the atomic parameters, except in so far as the atoms are fixed in special positions by the symmetry of the unit cell, though some general conclusions have been drawn from the experimental data. The only detailed study of parameters is the work of Náray-Szabo on $CaTiO_3$, which is discussed later.

In table 1 is given a list of all the compounds $A^{2+}B^{4+}O_3$ of perovskite type for which information is available. Some of these are known to be dimorphous, and it seems probable that others may be. Thus, $CdTiO_3$ has the ilmenite structure when fired below 1050°, the perovskite structure when fired above this; $BaTiO_3$ is tetragonal at room temperature but changes reversibly to cubic at about 120°, and also shows the rhombohedral form described later. It is possible that

the conflicting reports on $PbTiO_3$ are due to the occurrence of two different structures. This polymorphism should not be surprising in view of the frequency with which it occurs among minerals which have been thoroughly investigated, notably TiO_2 with its three forms, rutile, anatase and brookite.

Table 1.   Classification of compounds of perovskite type

Cubic	Tetragonal	Orthorhombic or monoclinic	Rhombohedral	Non-cubic, type not known
$SrTiO_3$ (G, H, M)	$BaTiO_3$ Iα (M)	$CaTiO_3$ (B, N, M)	$BaTiO_3$ II (M)	$PbZrO_3$ (N, H)
$SrZrO_3$ (H, M)	$PbTiO_3$ (N, M, R)	$CaZrO_3$ (RES, M)		$PbSnO_3$ (N)
$SrSnO_3$ (H, M)	$PbZrO_3$ (M)	$PbTiO_3$ (CE)		$SrCeO_3$ (H)
$SrHfO_3$ (H)		$CdTiO_3$ (M)		$CdSnO_3$ (N)
$BaZrO_3$ (H, N, M)		$CaSnO_3$ (R, M)		
$BaSnO_3$ (N, M)		$CaCeO_3$ (N)		
$BaCeO_3$ (H)		$CdCeO_3$ (N)		
$BaPrO_3$ (H)		$PbCeO_3$ (N)		
$BaThO_3$ (H, M)				
$BaTiO_3$ Iβ (M)				

Letters in brackets indicate the authors whose classification is quoted:

G = Goldschmidt, 1927.	R = Rooksby, 1945.
H = Hoffmann, 1935.	B = Barth, 1925.
N = Naray-Szabo, 1943.	CE = Cole and Espenschied, 1937.
M = Megaw, 1946 (present work).	RES = Ruff, Ebert and Stephan, 1929.

§3. EXPERIMENTAL METHOD

All the substances investigated in the present work were prepared by firing together stoichiometric mixtures of the appropriate oxides or carbonates, at temperatures specified in each case. No special precautions were taken about quenching.

For the x-ray work, specimens were prepared either by rolling rods from a paste of the material mixed with powdered gum tragacanth, or by filling small celluloid tubes (Lonsdale and Smith, 1941). In general, powder photographs were taken with filtered Cu Kα radiation in a 19-cm. camera; for some of the cubic materials, a 9-cm. camera or unfiltered radiation was used. With a Philips sealed tube, run at 30 kv. and 20 ma., exposures in the 19-cm camera were from 6 to 15 hours. The photographs were measured with a glass scale, and the spacings calculated by the method of Bradley and Jay (1932). No special precautions were taken to allow for absorption, and the spacings derived from low-angle lines can therefore not be taken as accurate; but they were sufficiently good to index the high-angle lines, and these, using the $cos^2\theta$ extrapolation, gave very exact values of the cell dimensions. For the cubic structure this extrapolation was straightforward; for the others, a method of successive approximations was used, as described below.

Estimates of the intensities were made visually.

§4. INTERPRETATION OF THE STRUCTURES

Photographs of typical substances are reproduced in figure 2. It will be seen from this that for most of the compounds the main lines (at least at the

Figure 2.

Powder photographs taken in a 19-cm. camera with Cu Kα radiation.
N values refer to pseudo-cell; lines requiring doubled cell are indicated by fractional values of N.
* The dots between 5 and 10 are all slightly displaced to the right,

N	Cubic indices	Tetragonal with $c/a>1$	Rhombohedral $\alpha>90°$	Monoclinic with $a^*=b^*=c^*$
1	100	001, 100+010 Doublet with outer line stronger		100+001+010 Single line
2	110	101+011, 110 Doublet with inner line stronger	Symmetrical doublet	101, 110+011+$1\bar{1}0$+$01\bar{1}$, $10\bar{1}$ Symmetrical triplet; middle line strongest
3	111	111 Single line	$11\bar{1}+\bar{1}11+\bar{1}11$, 111 Doublet with inner line stronger	$11\bar{1}$, 111 Symmetrical doublet
4	200	↓	as 100	↑
5	210	102+012, 201+021, 210+120 Three lines of equal intensity; outer pair a very close doublet	$2\bar{1}0$, 210 Symmetrical doublet	$20\bar{1}+10\bar{2}$, 210+012+120+021+$2\bar{1}0$+$01\bar{2}$+$1\bar{2}0$+$02\bar{1}$, 201+102 Symmetrical triplet with middle line strongest
6	211	112, 211+121 Doublet with outer line stronger	$\bar{2}11$, $2\bar{1}1$, $2\bar{1}1$, 211 Unsymmetrical triplet with middle line strongest, the inner pair being a very close doublet	$21\bar{1}+11\bar{2}$, $12\bar{1}$, 121, 211+112 Symmetrical set of four lines, consisting of two close pairs; outermost and innermost lines stronger than middle two

Note—The descriptions of the line complexes refer to their appearance to visual examination only, and are not meant to be quantitatively exact.

low-angle end of the photograph) are simply related to the cubic pattern, a single line (or α-doublet) of the latter being replaced by a complex group.

The *hkl* indices of lines in any group are permutations of those characterizing the cubic line, which are known. A survey was made, for each kind of departure from cubic symmetry, of the possible permutations giving crystallographically distinct spacings in every group. From this the type of multiplet formation to be expected could be seen. Part of this survey is given in table 2. Comparison of the photographs with this table showed which type of structure was present; where the deformation was fairly large, the lowest-angle lines were used, but where it was smaller, sufficient resolution was only obtained at higher angles. Further refinement was achieved by the use of high-angle lines.

Some of the substances gave weaker lines intermediate in position between those that could be indexed from the small pseudo-cell as just described. These could be accounted for by doubling the cell-edges. To retain the pseudo-cubic description, all three cell-edges must be doubled ; whether this is in fact necessary, or whether the true cell may only need the doubling of one edge, can only be determined by an exact indexing of all the lines in question. This has not been carried out, owing to the lengthy calculations needed in some cases and the ambiguity which results from overlapping.

§5. GEOMETRY OF THE STRUCTURE AND EXACT EVALUATION OF CELL CONSTANTS

(i) *Cubic*

This structure is straightforward, and no discussion is needed.

(ii) *Tetragonal*

This structure is derived from the ideal cubic by a homogeneous extension or compression along one axis, which becomes the *c*-axis. In the substances investigated, there is no evidence of any doubling of the cell-edge; hence the atomic coordinates are as in the cubic structure. The general framework of linked octahedra round the smaller cations, with the larger cations in the holes, remains unaltered.

For a tetragonal structure, the spacings are given by:

$$d_{hkl} = \frac{1}{\sqrt{\left(\dfrac{h^2+k^2}{a^2} + \dfrac{l^2}{c^2}\right)}}$$

$$= \frac{a}{\sqrt{\left(h^2+k^2+\dfrac{l^2}{r^2}\right)}}, \qquad \qquad \ldots\ldots(1)$$

where $c/a = r$. Approximate values of c and a may be found by single measurements of 004, 400, hence giving r.

Let $\qquad \qquad h^2+k^2+\dfrac{l^2}{r^2} = N' = N+l^2\left(\dfrac{1}{r^2}-1\right)$.

(*Cf.* $h^2+k^2+l^2 = N$ for a cubic crystal.)

Let $\qquad \qquad r = 1+x$, where x is small.

Then $\qquad \qquad N' = N-2xl^2$

Hence, where x^2 is negligible,

$$a = d_{hkl}\sqrt{N'}_{hkl} = d_{hkl}(\sqrt{N})\left(1 - \frac{xl^2}{N}\right), \qquad \ldots\ldots(2)$$

or, more exactly,

$$a = d_{hkl}\sqrt{(N)}\bigg/\sqrt{\left[1 + \frac{l^2}{N}\left(\frac{1}{r^2} - 1\right)\right]}. \qquad \ldots\ldots(2a)$$

Considering lines of high angle only, suitable for a Bradley-Jay extrapolation, $\sqrt{N'}$ can be calculated from the known r, and hence values of a found which are plotted against $\sin^2\theta$ in the usual way. An example is shown in figure 3 (*a*). If
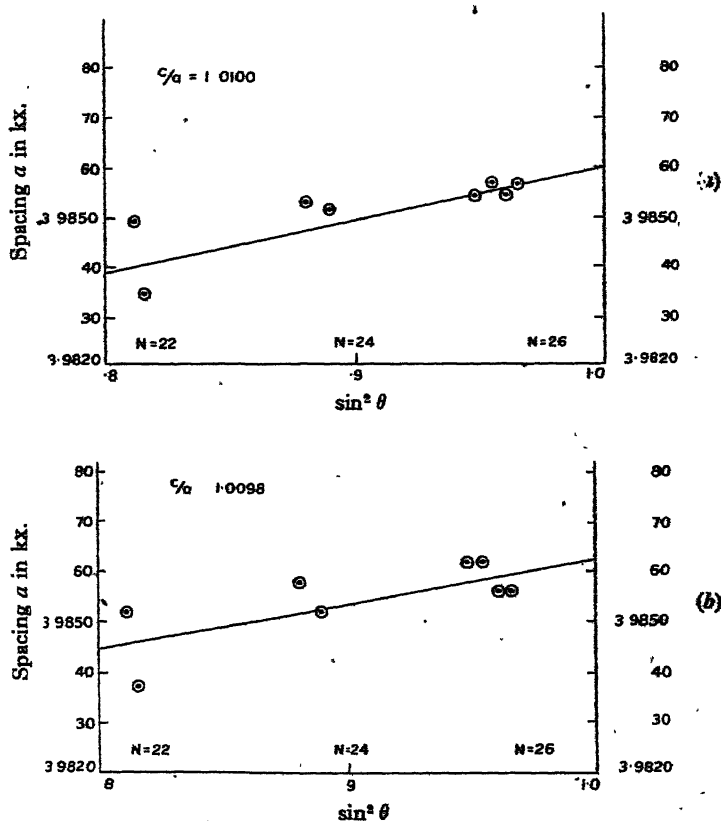
Figure 3. Extrapolation of a against $\sin^2\theta$ for BaTiO$_3$:
(*a*) using value of c/a finally adopted; (*b*) using incorrect value of c/a.
Reflections with $N=25$ are omitted as being less accurately measurable
owing to their weakness.

r is accurate, all the points will be on the same straight line; but if it is inaccurate, as in figure 3 (*b*), a line drawn through points corresponding to the complex with a given N will be inclined at an angle to the best line through all the points with different N's. The angle between these two lines is used to correct r, and a second approximation made, if necessary, to obtain the best value of a. The method is tedious for a single photograph, but can be systematized and shortened where many samples are being examined. In practice, while the above approximation (2) served for indexing the lines, the more exact expression (2a) was used for calculating the cell dimensions.

(iii) *Monoclinic and orthorhombic structures*

All the substances in this group that have so far. been investigated have a monoclinic pseudo-cell derived from the cubic cell by a simple shear, leaving the cell-edges a and c equal. Geometrically the lattice thus achieved has ortho-rhombic symmetry, and should be referred to new a and c axes, which are diagonals of the (010) face of the pseudo-cell, as shown in figure 4 (where the β-angle is exaggerated for clarity).

Some of the evidence on this point is considered later. It is, however, legitimate for purposes of calculation and comparison to retain the monoclinic pseudo-cell, leaving a transference to the orthorhombic axes to be made when required for a full determination of the structure.

There is also a doubling of the cell-edge, which implies change of atomic parameters, but these changes must be fairly small. This is considered in more detail later, in the discussion of the individual structures.
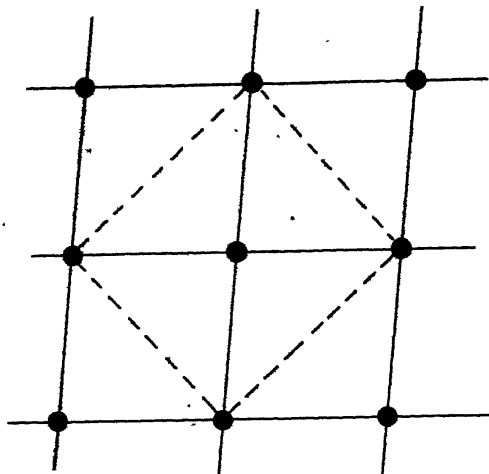
Figure 4. Projection of orthorhombic lattice on (010) plane, showing monoclinic pseudo-cells (full lines) and orthorhombic cell (dotted lines). The angle β has been made markedly different from 90° to show up the geometrical relationship.

The structure remains homotypic with those previously described; the network of linked octahedra is retained, though these are no longer regular nor symmetrically placed.

To find the cell dimensions, a method of approximations was used. For a monoclinic crystal,

$$d_{hkl} = (h^2 a^{*2} + k^2 b^{*2} + l^2 c^{*2} - 2 hl a^* c^* \cos \beta)^{-\frac{1}{2}}, \qquad \ldots\ldots (3)$$

where a^*, b^*, c^* are the reciprocal lattice dimensions:

$$a^* = 1/a \sin \beta, \quad b^* = 1/b, \quad c^* = 1/c \sin \beta. \qquad \ldots\ldots (4)$$

If $a^* = c^* = rb^*$ (in accordance with observation),

$$d_{hkl} = \frac{a \sin \beta}{\sqrt{\left(h^2 + \dfrac{k^2}{r^2} + l^2 - 2hl \cos \beta \right)}}.$$

Let
$$N'' = h^2 + \frac{k^2}{r^2} + l^2$$

$$= h^2 + k^2 + l^2 + k^2\left(\frac{1}{r^2} - 1\right)$$

$$= N + k^2\left(\frac{1}{r^2} - 1\right),$$

and let $\qquad r = 1 + y$, where y is small. $\qquad\qquad \ldots\ldots(5)$

Then
$$N'' = N - 2yk^2.$$

$$\therefore \quad d_{hkl} = \frac{a \sin \beta}{\sqrt{(N - 2yk^2 - 2hl \cos \beta)}}$$

$$= \frac{a \sin \beta}{\sqrt{N}} \cdot \frac{1}{\left(1 - \frac{k^2 y}{N} - \frac{hl}{N} \cos \beta\right)} \qquad \ldots\ldots(6)$$

Hence for two planes hkl, $hk\bar{l}$, with a mean spacing d, we have

$$\frac{d_{hkl} - d_{hk\bar{l}}}{d} = -\frac{2hl}{N} \cos \beta. \qquad\qquad \ldots\ldots(7)$$

This can be used to calculate β. Again, comparing an $hk0$ line with the geometric mean of an $h0l$, $h0\bar{l}$ pair from the same complex, we have

$$d = (d_{h0l} \cdot d_{h0\bar{l}})^{1/2} = \frac{a \sin \beta}{\sqrt{N}}.$$

$$\therefore \quad d_{hk0} = d\left(1 + \frac{k^2}{N} y\right). \qquad\qquad \ldots\ldots(8)$$

This gives y.

With the best values of $\cos \beta$ and y so obtained, $a \sin \beta$ can be calculated from d_{hkl} for a number of high-angle lines, and the Bradley-Jay extrapolation carried out to get the correct values.

All the above work is done with main lines which can be indexed from the pseudo-cell, and in fact this indexing is used in tables 6 to 9. It remains to identify the extra lines. For this purpose the cell-dimensions are doubled, and spacings calculated from (6) for all possible values of hkl; comparison of these with the observed spacings serves to index the lines. In a good many cases the accuracy was not great enough to do this unambiguously, but it was sufficient to show that all lines could be indexed on the assumption of a doubled cell-edge. *

In practice it was found for all these structures that y was small compared with $\cos \beta$. This can be seen from the photo of $CdTiO_3$ shown in figure 2 (f); if y were zero, as assumed in table 2, line 8 would be a (visually) symmetrical triplet, and the actual displacement of the middle line towards the low-angle side indicated a small positive value of y. In $CaSnO_3$, the displacement was towards the high-angle side, indicating a negative value. In $CaTiO_3$ and $CaZrO_3$, the displacement was too small to be observed at these angles, but a small negative value of y was required to explain the high-angle lines.

---

* In this connection it may be noted that in Rooksby's photograph of $CaSnO_3$, of the four lines marked 111, only the two central ones belong to this complex, the outer pair being extra lines requiring a doubled cell, referred to which their indices are $13\bar{1} + 131$ and $320 + 230$.

(iv) *Rhombohedral structure*

A rhombohedral perovskite structure, not previously described, has been found with $BaTiO_3$. Here the distortion is produced by a compression along a triad axis of the ideal cubic cell. There are extra lines not yet fully interpreted, and the true cell is almost certainly larger than the pseudo-cell. The photograph is reproduced in figure 2 (*g*).

For the rhombohedral structure, we may use a simplified expression for the spacing, since the observed departure from cubic symmetry is so small.

If α is the internal angle between the rhombohedral axes, we may neglect terms in $\cos^2 \alpha$ and $\cos^3 \alpha$ and write

$$d_{hkl} = \left(\frac{h^2 + k^2 + l^2}{a^2} - 2 \frac{hk + kl + lh}{a^2} \cos \alpha \right)^{-\frac{1}{2}}$$

$$= \frac{a}{\sqrt{N}} \left(1 + \frac{hk + kl + lh}{N} \cos \alpha \right). \qquad \ldots \ldots (9)$$

Hence
$$\frac{d_{hkl} - d_{hkl}}{d} = \frac{2}{N}(h+k)\, l \cos \alpha, \qquad \ldots \ldots (10)$$

whence α can be found.

Experimental results

Complete tables of spacings, observed and calculated, will not be published, but lists have been sent for inclusion in the A.S.T.M. index. Certain spacings used in determining cell dimensions are shown in tables 3 to 9. The values so obtained are given in tables 10 to 13.

Table 3. Barium titanium oxide

N	Indices	Intensity	θ (Cu $K\alpha_1$ or α_2)	log *d*	*a* (kx.) ($c/a = 1 \cdot 0100 \pm 0 \cdot 0001$)
22	323 α_1	s	64·309	$\bar{1}$·93096	3·9849
	332 β_1	s	64·620	$\bar{1}$·92884	3·9835
	332 α_2	ms	64·931	$\bar{1}$·92981	3·9832
24	224 α_1	ms	69·838	$\bar{1}$·91322	3·9853
	224 α_2	m	70·228	$\bar{1}$·91323	3·9853
	422 α_1	s	70·636	$\bar{1}$·91105	3·9851
	422 α_2	ms	71·074	$\bar{1}$·91098	3·9844
25	304 α_1	w	73·430	$\bar{1}$·90418	3·9847
	403 α_1	w	73·928	$\bar{1}$·90305	3·9855
	430 ⎫ 500 ⎬ α_1	w	74·612	$\bar{1}$·90161	3·9864
	,, α_2	vw	75·203	$\bar{1}$·90148	3·9852
26	105 α_1	ms	76·956	$\bar{1}$·89714	3·9854
	314 α_1	s	77·805	$\bar{1}$·89567	3·9857
	413 α_1	s	78·565	$\bar{1}$·89446	3·9854
	431 ⎱ 501 ⎰ α_1	vs	79·441	$\bar{1}$·89317	3·9856
	431 ⎱ 501 ⎰ α_2	s	80·276	$\bar{1}$·89312	3·9852

Extrapolated, $a = 3\cdot9860 \pm 0\cdot0500$

Table 4. Lead titanium oxide

N	Indices	Intensity	θ (Cu Ka_1)	log d	a (kx.) ($c/a = 1.064 \pm 0.001$)
24	422	s	73·38	$\bar{1}$·90463	3·8947
25	403	mw	75·18	$\bar{1}$·90079	3·8944
27	115	w	75·83	$\bar{1}$·90001	3·8982
26	314	m	76·04	$\bar{1}$·89913	3·8944
26	413	s	80·56	$\bar{1}$·89201	3·8953
25	430 500 }	w	80·74	$\bar{1}$·89181	3·8974

Extrapolated, $a = 3.896_5 \pm 0.001$

Table 5. Lead zirconium oxide

N	Indices	Intensity	θ (Cu Ka_1)	log d	a (kx.) ($c/a = 0.988 \pm 0.001$)
24	422	m	65·53	$\bar{1}$·92662	4·1458
26	501 431 }	ms	70·97	$\bar{1}$·91017	4·1482
	413	m	71·68	$\bar{1}$·90835	4·1463
	314	ms	72·18	$\bar{1}$·90711	4·1480

Extrapolated, $a = 4.150_2 \pm 0.001$

Table 6. Calcium titanium oxide

N	Indices (using pseudo-cell)	Intensity	θ (Cu Ka_1)	log d	$a \sin \beta$ (kx.) $\beta = 90° 40'$ $\gamma = -0.0012$
20	40$\bar{2}$+20$\bar{4}$	mw	63·77	$\bar{1}$·93297	3·8149
	420+024 }	s	64·46	$\bar{1}$·93039	3·8122
	240+042 }				
	402+204	ms	64·81	$\bar{1}$·92913	3·8156
22	32$\bar{3}$	w	70·16	$\bar{1}$·91232	3·8149
	33$\bar{2}$+23$\bar{3}$	m	70·44	$\bar{1}$·91159	3·8162
	33$\bar{2}$+23$\bar{3}$	ms	71·45	$\bar{1}$·90893	3·8165
			71·65	$\bar{1}$·90842	3·8173
24	4$\bar{2}$2+$\bar{2}$24	ms	79·33	$\bar{1}$·89331	3·8181
	24$\bar{3}$	s	80·14	$\bar{1}$·89221	3·8174
	242	ms	81·54	$\bar{1}$·89052	3·8173
	422+224	m	82·01	$\bar{1}$·89001	3·8190

Extrapolated, $a \sin \beta = 3.818_3 \pm 0.0006$

Table 7. Calcium zirconium oxide

N	Indices (using pseudo-cell)	Intensity	θ (Cu Ka_1)	log d	$a \sin \beta$ (kx.) $\beta = 91°\ 43'$ $y = -0.0010$
24	$42\bar{2}+22\bar{4}$	m	68·88	$\bar{1}$·91595	3·9974
	$24\bar{2}$	m	69·71	$\bar{1}$·93156	3·9976
	242 $\left.\begin{array}{c}40\bar{3}+30\bar{4}\end{array}\right\}$	ms	71·32	$\bar{1}$·90927	$\left\{\begin{array}{l}3·9981\\3·9989\end{array}\right.$
	$422+224$	ms	72·04	$\bar{1}$·90743	3·9990
25	$\left.\begin{array}{c}430+034\\340+043\end{array}\right\}$	s	74·05	$\bar{1}$·90281	$\left\{\begin{array}{l}3·9990\\4·0000\end{array}\right.$
26	$41\bar{3}+31\bar{4}$	m	75·13	$\bar{1}$·90055	3·9995
	$\left.\begin{array}{c}50\bar{1}+10\bar{5}\\403+304\end{array}\right\}$	w	77·02	$\bar{1}$·89701	$\left\{\begin{array}{l}3·9992\\4·0010\end{array}\right.$
	$43\bar{1}+13\bar{4}$	m	77·36	$\bar{1}$·89643	3·9999
	$34\bar{1}+14\bar{3}$	ms	77·73	$\bar{1}$·89579	3·9992
	$510+015$	ms	78·48	$\bar{1}$·89459	4·0002
	$150+051$	ms	78·77	$\bar{1}$·89416	4·0002
	$341+143$	ms	79·64	$\bar{1}$·89291	4·0011
	$431+134$	ms	79·93	$\bar{1}$·89250	4·0005
	$413+314$	s	83·28	$\bar{1}$·88875	4·0017

Extrapolated, $a \sin \beta = 4·0015 \pm 0·0015$

Table 8. Calcium tin oxide

N	Indices (using pseudo-cell)	Intensity	θ (Cu Ka_1)	log d	$a \sin \beta$ (kx.) $\beta = 91°\ 30'$ $y = -0.0024$
20	$40\bar{2}+20\bar{4}$	w	59·74	$\bar{1}$·94937	3·9382
	$420+024$	ms	60·84	$\bar{1}$·94451	3·9377
	$240+042$	ms	60·97	$\bar{1}$·94408	3·9392
	$402+204$	mw	61·90	$\bar{1}$·94023	3·9380
22	$32\bar{2}$	m	64·93	$\bar{1}$·92872	3·9394
	$33\bar{2}+23\bar{3}$	ms	65·42	$\bar{1}$·92703	3·9405
	$332+233$	s	67·30	$\bar{1}$·92078	3·9405
	323	s	67·71	$\bar{1}$·91949	3·9399
24	$42\bar{2}+22\bar{4}$	m	71·32	$\bar{1}$·90927	3·9423
	$24\bar{2}$	m	72·36	$\bar{1}$·90668	3·9406
	242	mw	74·02	$\bar{1}$·90288	3·9408
25 24	$30\bar{4}+40\bar{3}$	ms	74·30	$\bar{1}$·90227	3·9421
	$422+224$	ms	74·63	$\bar{1}$·90157	3·9409
25	$\left.\begin{array}{c}034+430\\043+340\end{array}\right\}$	s	77·45	$\bar{1}$·89626	$\left\{\begin{array}{l}3·9411\\3·9435\end{array}\right.$
26 25	$31\bar{4}+41\bar{3}$	ms	79·10	$\bar{1}$·89367	3·9438
	$304+403$	m	80·84	$\bar{1}$·89133	3·9422
26	$10\bar{5}+50\bar{1}$	m	81·63	$\bar{1}$·89041	3·9420
	$13\bar{4}+43\bar{1}$	ms	82·35	$\bar{1}$·88964	3·9422
	$14\bar{3}+34\bar{1}$	ms	83·09	$\bar{1}$·88893	3·9425

Extrapolated, $a \sin \beta = 3·9437 \pm 0·001$

Table 9.   Cadmium titanium oxide

N	Indices (using pseudo-cell)	Intensity	θ (Cu $K\alpha_1$)	log d	$a \sin \beta$ (kx.) $\beta = 91°\ 10'$ $y = +0.0044$
18	41$\bar{1}$+11$\bar{4}$ } 14$\bar{1}$	m	59·15	$\bar{1}$·95201	{ 3·7809 3·7797
	141 330+033 }	ms	59·39	$\bar{1}$·95096	{ 3·7789 3·7812
	411+114	m	60·06	$\bar{1}$·94794	3·7795
20	240+042	s	64·97	$\bar{1}$·92860	3·7809
	420+024	s	65·24	$\bar{1}$·92766	3·7825
	402+204	m	66·48	$\bar{1}$·92342	3·7799
22	33$\bar{2}$+$\bar{2}$3$\bar{3}$	m	71·15	$\bar{1}$·90790	3·7809
	332+233	m	73·11	$\bar{1}$·90489	3·7815
	323	ms	73·81	$\bar{1}$·90332	3·7825
24	42$\bar{2}$+22$\bar{4}$ } 24$\bar{2}$	m	81·35	$\bar{1}$·89073	{ 3·7806 3·7856
	242	m	84·72	$\bar{1}$·88761	3·7838

Extrapolated, $a \sin \beta = 3.7833 \pm 0.001$

Table 10.   Cubic structures

Compound	Firing temperature (°c.)	Cell-edge a_0 (kx.)
$SrTiO_3$	1250	3·897$_9$ ± 0·001
$SrZrO_3$	1250	4·093 ± 0·001
$SrSnO_3$	1380	4·0254 ± 0·0005
$BaZrO_3$	1280	4·181$_5$ ± 0·001
$BaSnO_3$	1300	4·108$_5$ ± 0·001$_5$

Table 11.   Tetragonal structures

Compound	Firing temp. °c.	Cell-dimensions a (kx.) c (kx.)		c/a	(Vol.)$^{1/3}$ (kx.)
Barium titanium oxide	1300	3·986$_0$	4·0259	1·0100±0·0001	3·9993±0·0005
Lead titanium oxide	1180	3·896$_0$	4·144$_0$	1·063$_5$±0·001	3·998 ±0·001
Lead zirconium oxide	1250	4·150$_3$	4·100	0·988 ±0·001	4·133 ±0·001

Helen D. Megaw

Table 12. Orthorhombic structures

Compound		Calcium titanium oxide	Calcium zireonium oxide	Calcium tin oxide	Cadmium titanium oxide
Firing temp. (°C.)		1500	1500	1400	1200
Monoclinic pseudo-cell	$\cos\beta$	-0.0115 ± 0.0003	-0.0300 ± 0.0010	-0.0262 ± 0.0005	-0.0204 ± 0.0006
	y	-0.0012 ± 0.0003	-0.0010 ± 0.0005	-0.0024 ± 0.0004	$+0.0044\pm0.0006$
	$a\sin\beta$	$3.818_8\pm0.001$	$4.001_5\pm0.001_5$	$3.942_7\pm0.001$	$3.783_8\pm0.001$
	$a=c$	3.819	4.003	3.944	3.784
	b	3.815	3.997	3.933	3.800
	β	90° 40′±1′	91° 43′±4′	91° 30′±2′	91° 10′±2′
	$a/b=c/b$	1.0011	1.0015	1.0028	0.9958
	$(vol.)^{1/3}$	3.818	4.001	3.940	3.789
Orthorhombic large cell	A	10.740	11.152	11.034	10.594
	B	7.630	7.994	7.866	7.600
	C	10.864	11.492	11.302	10.812
	A/B	1.4076	1.3951	1.4028	1.3939
	C/B	1.4239	1.4376	1.4367	1.4226

Table 13. Rhombohedral structure

Barium titanium oxide II

$$a=b=c=4.035\pm0.005 \text{ kx.}$$

$$\alpha=90° 19'\pm3'$$

$$(vol.)^{1/3}=4.035 \text{ kx.}$$

§6. DISCUSSION OF RESULTS

Cubic

These need no further comment.

Tetragonal

There are considerable discrepancies between the accounts in the literature of the three substances.

$BaTiO_3$ was said by Naray-Szabo to be strictly cubic, though Goldschmidt classed it as only pseudo-cubic, but without further detail. It has been found in the present work that while it is normally tetragonal, polymorphous forms (such as the rhombohedral one, of which a preliminary account is given here) do occur under exceptional conditions, and these are at present being investigated. Only structures which exist at room temperature are considered here; it is hoped to deal with temperature changes in a subsequent paper.

The table gives cell-dimensions of typical material, but these are not necessarily the same for all conditions of preparation. In most of the material investigated, only small variations of axial ratio occur, leaving the cell-volume constant.

This leaves unexplained the discrepancy between Rooksby's values of the cell-dimensions and those found here, but it may perhaps be due to the polymorphism of the substance.

PbTiO$_3$ was originally said by Naray-Szabo (1943a) to belong to the same sub-type as CaTiO$_3$ (see below) ; but in a later paper (1943b) he showed it to be tetragonal, with cell-dimensions substantially in agreement with those found here, and space-group D$^1_{4h}$ – P4/mm.   Cole and Espenschied found it orthorhombic, with $a = 4·000, b = 4·211, c = 3·875$, $a:b:c = 0·95 : 1 : 0·92$, and moreover, found that it was the only intermediate compound occurring in the system PbO-TiO$_2$.   PbZrO$_3$ was classed by Naray-Szabo with CaTiO$_3$. The material investigated in the present work gives weak lines besides those due to the tetragonal structure; these have not been interpreted, and may be due to other compounds formed, to a second structure simultaneously present, or to a doubling of one or more of the cell-edges (in which case the dimensions given refer to the pseudo-cell).

Rhombohedral structure

This was found in a preparation of BaTiO$_3$ fired at a high temperature. Further work is in progress, and the report here is only a preliminary note, which seems desirable in view of the novelty among double oxides of the geometrical type illustrated.   A rhombohedral perovskite-type structure has been described by Naray-Szabo (1943b) for ScF$_3$, with $a = 4·022$ kx. and $\alpha = 89°\ 34\frac{1}{2}'$, and comparison should also be made with NiO, found by Rooksby (1943) to be rhombohedral with a very slight departure from cubic symmetry.   The amount of departure from cubic symmetry in rhombohedral BaTiO$_3$, as in the tetragonal form, seems to vary with conditions of preparation, and here takes the form of a variation in the interaxial angle α.   The value quoted in the table is the largest observed; against this, another preparation gave a strictly cubic structure with about the same cell-volume, which is appreciably larger than that of the tetragonal form.

Monoclinic and orthorhombic structures

Of the structures tabulated, CaTiO$_3$ has been investigated by Naray-Szabo (1943a), and CaZrO$_3$ by Ruff, Ebert and Stephan (1929). Their results agree with those in the present work within the limits of accuracy claimed. It is possible that small variations in the cell-dimensions of different samples may exist, as in barium titanate; such a variation is stated by Goldschmidt and Rait (1943) to occur when CaTiO$_3$ takes up CaO in solid solution.

Naray-Szabo, using single-crystal methods, determined the atomic parameters of CaTiO$_3$. He found that the intensities are consistent with the assumption that the cations are not displaced, while the oxygens are displaced in such a way as to retain approximately regular octahedra round the Ti's, though leaving them tilted relative to one another and making a puckered network. His structure is shown in figure 5.

The displacement of the atoms from their special positions must be consistent with their true symmetry. Naray-Szabo's set of oxygen parameters has monoclinic symmetry only; however, very slight changes in some of them would give orthorhombic symmetry with the new axes postulated above, and it is probable that such changes are within his limits of accuracy.

It is noteworthy that in all the compounds of this type that have been examined, the a and c edges of the pseudo-cell remain equal. No difference is recorded by any author, and none was found in the present work, though special attention was paid to the point and the technique was such that very small differences could have been detected. While the evidence is not final, an orthorhombic structure thus seems much more probable, the new [001] and [100] axes being the [101] and [10$\bar{1}$] directions of the monoclinic lattice.

It was mentioned earlier that the true cell required the doubling of some of the edges of the pseudo-cell, but not necessarily all. It is simplest, however, to make the provisional assumption that all are doubled, transfer to the ortho-rhombic axes, and then look for special absences among the indices of reflexions. This has not been carried out in detail; enough lines have been indexed to confirm the fact that all can be indexed on this large cell, but the overlaps are too numerous
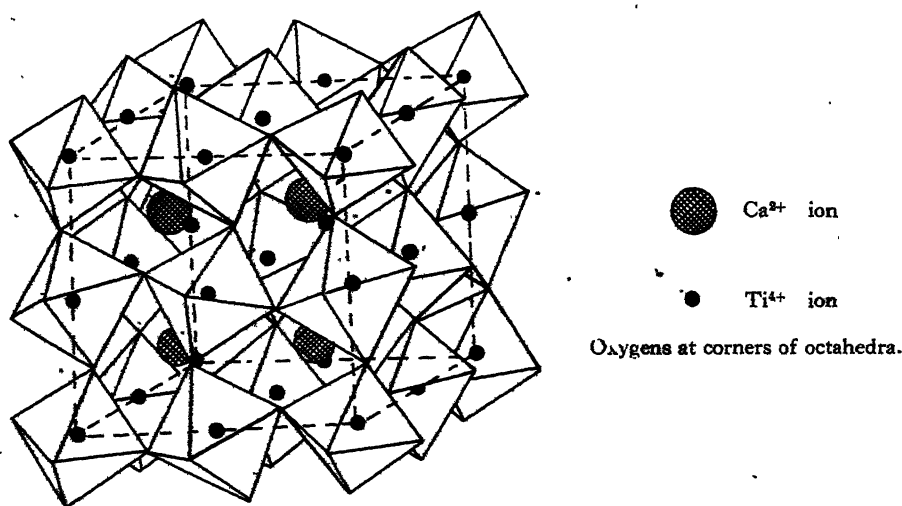
Figure 5. Structure of $CaTiO_3$ (according to Náray-Szabó).

for completely unambiguous indexing. The available evidence suggests that the doubled value of b is necessary, and also that of either a or c, possibly both. The volume of the true cell is then either $8a^3$ or $16a^3$ approximately.

While the extra lines, needing the doubled cell, are numerous and of moderate intensity for $CaTiO_3$, they are very few and weak for $CaZrO_3$ (as can be seen by comparison of figures 2(d) and (e)), $CaSnO_3$ occupying an intermediate position. This strongly suggests that they are due to displacements either of the oxygens alone or of oxygens and Ca ions, leaving the 4-valent cation, whose contribution to the intensity increases rapidly in the order Ti^{4+}, Sn^{4+}, Zr^{4+}, undisplaced from the special position it occupied in the simple cell. This agrees with the conclusion reached by Náray-Szabó, from detailed work on the atomic parameters in $CaTiO_3$, that both cations remain in special positions and only the oxygens are displaced.

By contrast, there are certain striking intensity relationships in $CdTiO_3$. Firstly, the lines needing the doubled cell are prominent at high angles, where they are of the same order of intensity as the main lines (see figure 2(f)). This implies displacement of the cations from their special positions, as the oxygen

contribution could not by itself effect this.   Both from the intensity, and because the Cd ion is held by weaker electrostatic bond strengths in the structure (2/12 for the Cd-O bond as compared with 4/6 for Ti-O), it seems likely that it is the Cd ions that are displaced.   Secondly, the reduction in intensity of "main" lines with high values of orthorhombic *l* index, (that is, with monoclinic *h–l* large), and the preponderance with appreciable intensity of "doubled-cell" lines also with high orthorhombic *l*, independently suggest that the atomic displacement is along the orthorhombic *z*-axis, that is, along the long diagonal of the monoclinic pseudo-cell.   Even without any displacement of the oxygens, this brings the Cd ion closer to six of its neighbouring oxygens and further away from the other six, thus making a first step towards the attainment of 6-coordination which is appropriate to the radius of the Cd ions.   Displacements of the oxygens are likely to reinforce this tendency.

§7. CRITERION FOR FORMATION OF DIFFERENT STRUCTURES.

Goldschmidt (1937) showed empirically that structures having the generalized perovskite type occur when the tolerance factor *t*, defined as

$$t = \frac{R_A + R_O}{\sqrt{2}(R_B + R_O)}$$

(where R_A, R_B, R_O are the ionic radii of the ions A, B, O respectively), lies between 0·80 and 1·00.   Lower values of *t* are associated with the corundum–ilmenite type (a rhombohedral structure).   Later authors showed that for values of *t* at the lower end of the perovskite range "deformed" (i.e. monoclinic or orthorhombic) structures occur.

Table 14.   Values of tolerance factor *t*

R_A (kx.) 6-coord. 12-coord.		R_B (kx.)	Ti⁴⁺	Sn⁴⁺	Zr⁴⁺	Th⁴⁺
			0·64	0·74	0·77	1·10
Ca²⁺	1·06	1·16	0·89	0·85	0·84	0·72
Sr²⁺	1·27	1·37	0·97	0·92	0·91	0·78
Ba²⁺	1·43	1·52	1·02	0·97	0·96	0·83
Pb²⁺	1·32	1·40	0·98	0·92	0·93	0·79
Cd²⁺	1·03	1·11	0·88	0·83	0·82	0·71

Table 14 shows the values of *t* for the compounds under discussion.   They are calculated from Goldschmidt's ionic radii except that Naray-Szabo's corrected value of 0·77 kx. has been accepted for Zr⁴⁺ ; this value has been confirmed by the present work, as appears in table 15.   A correction for 12-fold coordination has been applied to the R_A, as Goldschmidt's radii are quoted for 6-coordination; the correcting factor was taken from Pauling's *Nature of the Chemical Bond* (p. 368).

It should be noted that the Goldschmidt radii are valid only to an accuracy of 1 or 2%; they depend on the assumption that ions can be treated as rigid spheres, which is only true to this degree of accuracy, even when the appropriate corrections for coordination number have been made.

Comparison of table 14 with table 15 shows general agreement with the empirical rule; all the compounds observed to have the perovskite structure have t-values not less than 0·80, and those with the orthorhombic modification lie between 0·80 and 0·89. The t value of $BaThO_3$ seems rather low for its cubic structure, but this is probable due to an incorrect Th^{4+} radius (as explained later). The lead compounds are slightly anomalous. $BaTiO_3$ is unique in having a value of t above unity.

The explanation of these relationships may be better seen by consideration of the interionic distances A-O and B-O actually observed. These are given in

Table 15. Different structures observed

	Titanium	Tin	Zirconium	Thorium
Calcium	Orthorhombic * (referred to monoclinic axes, $\beta=90°\ 40'$)	Orthorhombic * (referred to monoclinic axes, $\beta=91°\ 30'$)	Orthorhombic * (pseudo-monoclinic with $\beta=91°\ 34'$)	—
Strontium	Cubic	Cubic	Cubic	—
Barium	Tetragonal $c/a=1·01$	Cubic	Cubic	Cubic
Lead	Tetragonal $c/a=1·06$	—	Tetragonal $c/a=0·99$	—
Cadmium	Orthorhombic * (referred to monoclinic axes, $\beta=91°\ 10'$)	—	—	—

* For relation of orthorhombic and monoclinic axes, see text.

table 16, with the sum of the Goldschmidt radii quoted for comparison. The figures in brackets for the orthorhombic structures are approximations, based on the cubic oxygen parameters. For $BaTiO_3$, figures are also given calculated from Naray-Szabo's parameters (which are probably partly derived from the assumption of Goldschmidt's distance). Owing to the low symmetry there are sets of A-O and B-O distances for each substance, and the figures give an idea of the range included.

To realize the significance of these figures, it must be understood that the dimensions of the structure are primarily determined by the B-O octahedra, where the strong valence bonds of value 2/3 keep the octahedra as nearly rigid units (though they may be tilted relative to each other). The network is only stable if the holes are occupied by A cations to restore electrical neutrality. Since the valence bonds to these are weak, of value 1/6, the size of A has much less effect

on the structure, but it must be small enough to fit in the hole and large enough to occupy it nearly completely. From table 16, the reasons for the occurrence of the different structures can now be explained in a very general way:—

(*a*) In all the cubic structures, the B-O distance is very little *less* than the sum of the radii, the difference not exceeding 2%. (An apparent exception is $BaThO_3$,

Table 16. Interionic distances (in kx.)

(*a*) A-O distances. Values for the same ion A derived from different compounds are read horizontally.

	Titanium	Tin	Zirconium	Thorium	Goldschmidt's $R_A + R_O$ (R_A corrected for 12-coordination)
Calcium	(2·70) (2·3–3·0)	(2·79)	(2·83)	—	2·48
Strontium	2·755	2·845	2·894	—	2·69
Barium	2·818 2·832	2·907	2·957	3·17	2·84
Lead	2·755 2·844	—	2·917 2·934	—	2·72
Cadmium	(2·69)	—	—	—	2·45

(*b*) B-O distances. Values for the same ion B derived from different compounds are read vertically.

	Titanium	Tin	Zirconium	Thorium
Calcium	(1·91) 1·94–1·96	(1·97)	(2·00)	—
Strontium	1·949	2·012	2·047	—
Barium	1·992 2·013	2·054	2·091	2·243
Lead	1·948 2·072	—	2·050 2·075	—
Cadmium	(1·90)	—	—	—
Goldschmidt's $R_B + R_O$	1·96	2·06	2·09	2·43

but the discrepancy may be taken as evidence against the accepted value of the Th^{4+} radius.) In the same structures, the A-O distance is *greater* than the sum of the radii by amounts up to 10%. These facts fit the above conditions postulated for the ideal structure.

(*b*) In the orthorhombic structures, if the oxygens remained in their positions of cubic symmetry and the actual cell dimensions were retained, the B-O distances would be too small; a uniform expansion, to give the correct B-O distances, would leave the A-O distances unduly large. The actual structure retains the normal B-O distances. but decreases the average 'A-O distance (and the cell dimensions) by tilting the octahedra, thus abandoning cubic symmetry. In $CdTiO_3$ the displacement of the Cd ions achieves the same effect.

(*c*) Barium titanate deviates from the normal in the opposite direction. The A-O distance is *less* than the sum of the radii, the B-O distance *greater*. In other words, the Ba ion is too big for the space available in the normal Ti-O network, and a compromise is achieved with some extension of the Ti-O network and some compression of the Ba ion. Why the deformation should be tetragonal is not immediately clear, but it seems possible that the abnormal volume available to the Ti ion may result in the formation of directed bonds.* It is known that the interactions of atoms and ions are very sensitive to volume changes.

(*d*) The lead compounds are anomalous. The A-O distances are less than the sum of the radii; hence one would expect that the B-O distances should be equal to or slightly less than the sum of the radii, and that the symmetry should be cubic. In fact one Ti-O distance is greater than the sum of the radii, and both structures are tetragonal. The explanation may lie in the character of the Pb^{2+} ion, in which the inner shells affect the valency; such an ion cannot be treated, except very roughly, as an undeformable sphere, and formation of directed bonds is very probable, predominating over the purely size effects which determine the other structures.

These explanations of the structure differences are very general, but they and the experimental results on which they are based may suggest the lines along which further theoretical work could be pursued.

REFERENCES

BARTH, T., 1925. *Norsk. Geol. Tidsskr.* 8, 201.
BRADLEY, A. J. and JAY, A. H., 1932. *Proc. Phys. Soc.* 44, 563.
COLE, S. S. and ESPENSCHIED, H., 1937. *J. Phys. Chem.* 41, 445.
EVANS, R. C., 1939. *Crystal Chemistry* (Cambridge : Univ. Press), p. 166.
GOLDSCHMIDT, H. J. and RAIT, J. R., 1943. *Nature, Lond.*, 152, 356.
GOLDSCHMIDT, V. M., 1927. *Geochem. Vert. d. Elem.* VII, VIII.
HOFFMANN, A., 1935. *Z. phys. Chem.* B, 28, 65.
LONSDALE, K. and SMITH, H., 1941. *J. Sci. Instrum.* 18, 133.
MEGAW, H. D., 1945. *Nature, Lond.*, 155, 484.
NARAY-SZABO, S., 1943 a. *Naturwissenschaften*, 31, 203 ; 1943 b. *Ibid.* 466.
ROOKSBY, H. P., 1943. *Nature, Lond.*, 152, 304 ; 1945. *Ibid.* 155, 484.
RUFF, O., EBERT, F. and STEPHAN, E., 1929. *Z. anorg. Chem.* 180, 215.

* It is hoped shortly to publish further work on this point.

THE MAGNETO-RESISTANCE OF HIGH COERCIVITY ALLOYS

By L. F. BATES,

University College, Nottingham

ABSTRACT. Measurements have been made of the change of resistance of ferromagnetic alloys of high coercivity when exposed to longitudinal and transverse magnetic fields. The changes in the cases of the permanent magnet materials alni, alnico (cast and sintered) and alcomax II are in marked contrast to those of pure ferromagnetic metals. The results are discussed on the basis of the domain concept in ferromagnetism.

§ 1. INTRODUCTION

SEVERAL new alloys of high coercivity and high $(BH)_{max}$ have become available for the manufacture of permanent magnets within recent years, but data on their magneto-resistance, i.e., the change of resistivity as a function of the magnitude and direction of an applied magnetic field, are not to hand. In the present communication an account is given of the measurement of the magneto-resistance of some representative materials, together with a discussion of the theoretical aspects of the data so obtained.

It is convenient to take a thin rod of the material whose resistance in the unmagnetized state is R_0, and measure the changes $\Delta R_{\parallel}$ and $\Delta R_{\perp}$ which respectively occur when a magnetic field H is applied parallel to and perpendicular to the long axis of the rod, and to plot the values of $\Delta R_{\parallel}/R_0$ and $\Delta R_{\perp}/R_0$ against H. This requires a definition of the unmagnetized state; for present purposes this is taken to be the state of the rod when, by means of a magnetic potentiometer, no difference of magnetic potential can be found over its length. Unfortunately, there is some ambiguity in the matter, since, for example, this result is obtained when the magnetic vectors are either all aligned parallel and antiparallel to the long axis of the rod, or are all aligned perpendicular thereto, in such a way that there is no resultant magnetization.

Important data for polycrystalline specimens of the ferromagnetic metals have hitherto been obtained by Bittel (1938), Englert (1932), Gerlach (1932), McKeehan (1930), Obata (1925), and Scharff (1935). A special study of the relation between the phenomena and the change of spontaneous magnetization as shown by the magnetocaloric effect at high temperatures was made by Potter (1931-32). In the case of permanent magnet materials, initial investigations must, of course, be confined to room-temperature measurements.

The former workers have shown that, for a pure specimen of a ferromagnetic metal, $\Delta R_{\parallel}/R_0$ normally shows a marked initial increase with the applied field; it rises to a maximum and eventually decreases linearly with the field, so that in very high fields it may become zero and, indeed, change sign. In the present work, a specimen of pure cobalt was used for comparison measurements, and the

upper curve of figure 1 shows the typical variation of $\Delta R_{\parallel}/R_0$. When the direction of H is perpendicular to the direction of the current in the specimen, the values of $\Delta R_{\perp}/R_0$ are at first subject to complications due to the high value of the demagnetization field, D.I., but when the demagnetization field is overcome, the value $\Delta R_{\perp}/R_0$ falls rapidly with increase in H and finally decreases with the applied field at the same rate as $\Delta R_{\parallel}/R_0$, as shown by the lower curve in figure 1. The two points nearest the origin on the lower curve lie above the H axis, but this is merely because the specimen is of finite cross-section and its long axis cannot be placed absolutely at right angles to the applied field; consequently, a small longitudinal component is always present.

The main features of the curves of figure 1 may be explained on the Weiss domain theory of ferromagnetism, if we suppose that magneto-resistance changes occur only when domain vectors turn through angles between 0 and 90° or when boundary displacements take place between neighbouring domains whose vectors
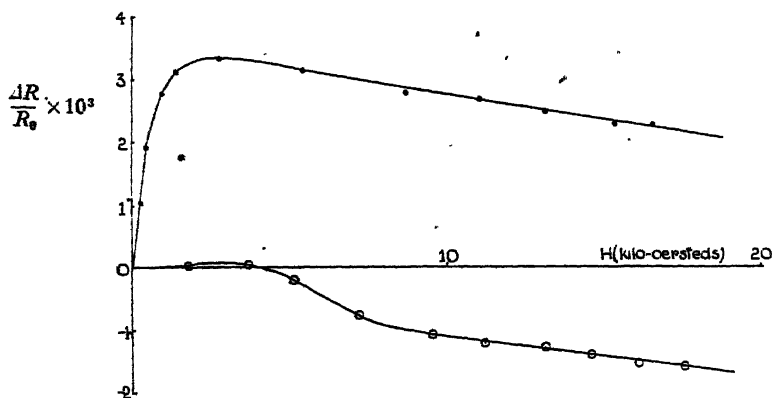
Figure 1. Cobalt.

● Longitudinal field.   ⊜ Transverse field.

are perpendicular to one another, and that they do not occur to any marked extent when vectors as, for example, those parallel to a direction of easy magnetization in a single crystal, turn through 180°. In order that Ohm's law may be satisfied, the resistance of a polycrystalline specimen of a ferromagnetic metal must be a function of $\overline{\cos^2\theta}$, where θ is the angle between the direction of the current in a portion of the specimen and that of the field when saturation conditions set in; the bar denotes the average taken over all portions of the specimen. Hence the resistivity of the material may in general be written as

$$\rho = A + B\overline{\cos^2\theta}, \qquad \ldots\ldots(1)$$

where A and B are constants. In an isotropic demagnetized state $\overline{\cos^2\theta}$ would have the value $\frac{1}{3}$, giving $\rho = \rho_0 = A + \frac{1}{3}B$. Under saturation conditions, with current and field parallel, we have

$$(\rho_{\parallel})_{\text{sat}} = A + B, \qquad \ldots\ldots(2)$$

and, with current perpendicular to the field,

$$(\rho_{\perp})_{\text{sat}} = A. \qquad \ldots\ldots(3)$$

Since, in practice, we measure relative changes of resistance only, it is more convenient to write for the general case of non-isotropic material, starting from $\rho_0 = A + B \overline{\cos^2 \theta_0}$,

$$\left(\frac{\Delta R_{\parallel}}{R_0}\right)_{\text{sat}} = \frac{B(1 - \overline{\cos^2 \theta})}{A + B \overline{\cos^2 \theta_0}}, \qquad \ldots\ldots(4)$$

$$\left(\frac{\Delta R_{\perp}}{R_0}\right)_{\text{sat}} = \frac{-B \overline{\cos^2 \theta_0}}{A + B \overline{\cos^2 \theta_0}} = a, \qquad \ldots\ldots(5)$$

and

$$\left(\frac{\Delta R_{\parallel}}{R_0}\right)_{\text{sat}} - \left(\frac{\Delta R_{\perp}}{R_0}\right)_{\text{sat}} = \frac{B}{A + B \overline{\cos^2 \theta}} = b, \qquad \ldots\ldots(6)$$

whence

$$\overline{\cos^2 \theta_0} = 1 \bigg/ \left[1 - \left(\frac{\Delta R_{\parallel}}{\Delta R_{\perp}}\right)_{\text{sat}}\right]. \qquad \ldots\ldots(7)$$

In the case of isotropic material, it follows that $(\Delta R_{\parallel})_{\text{sat}}$ should be equal to $-2(\Delta R_{\perp})_{\text{sat}}$, as found by Thiessen (1940) for annealed nickel, and, moreover, equation (7) means that $(\Delta R_{\parallel})_{\text{sat}}$ and $(\Delta R_{\perp})_{\text{sat}}$ must have opposite signs or $(\Delta R_{\parallel})_{\text{sat}}$ be zero. The value of b, which is the relative change in resistance when the magnetic state of the specimen is changed from transverse to longitudinal saturation, is clearly independent of the initial state of the specimen and, practically, of the value assumed for ρ_0. It should also be the same for polycrystalline as for single-crystal specimens. Positive values of b have been found for all materials previously investigated to date.

It is generally agreed that the intensity of magnetization of each Weiss domain may be increased slightly by the application of a strong field, and this provides the reason for the linear rates of change of $\Delta R_{\parallel}/R_0$ and $\Delta R_{\perp}/R_0$ at high values of the field, as depicted in figure 1. Theoretically, the two rates should be equal, and this is obviously the case for the cobalt specimen within the limits of experimental error. The values $(\Delta R_{\parallel}/R_0)_{\text{sat}}$ and $(\Delta R_{\perp}/R_0)_{\text{sat}}$ refer to saturation conditions extrapolated to zero external field. Consequently, to find the required value of $(R_{\parallel}/R_0)_{\text{sat}}$ from figure 1, the linear portion of the upper graph must be produced to cut the axis of ordinates, since we may neglect longitudinal demagnetization effects. The linear portion of the lower graph must be produced to cut the ordinate at $H = 2\pi I_s$, where I_s is the value of the intensity of magnetization, obtained by producing back the (I, H) curve at high field values to cut the I axis, and 2π is the demagnetization factor for a long narrow cylinder or rod magnetized at right angles to its long axis.

§2. EXPERIMENTAL PROCEDURE

Modern permanent magnet materials cannot, in general, be forged or machined. They are roughly cast to shape and ground to size. All specimens used in the present work were approximately 2·2 cm. long and 1 mm. square in cross-section. A potentiometer method was used to measure relative changes of resistance. The specimen was mounted for $\Delta R_{\parallel}$ measurements as follows. Copper end plates 0·8 mm. thick were screwed flush to the end faces of an ebonite block 1·9 cm. long and about 1·25 cm. by 1·0 cm. in cross-section. Current leads were soldered to the plates. The specimen was sunk into a saw-cut made in the common upper

surfaces of ebonite and copper plates, so that one side surface of the specimen was flush with the ebonite. Electrical connection between specimen and copper was made with a minimum quantity of soft solder and designed to give easily if the specimen suffered appreciable magnetostrictive change of dimensions. The potential leads consisted of thin tinned-copper wires inserted through fine holes drilled in the ebonite to lie with their axes slightly to one side of the specimen and some 1·8 cm. apart. The wires were pressed into contact with one side of the specimen by fine screws which ran parallel and close to the upper surface of the ebonite and at right angles to the axis of the specimen. The whole mounting was supported between the pole-faces of an electro-magnet, paper packing being used to prevent lateral motion and to allow for dimensional changes in the magnet, thermal expansion, etc.

The same ebonite and copper arrangement was used for measurements of ΔR_1, but experience showed the specimen to be severely strained unless special precautions were taken. Accordingly, the ebonite block was placed horizontally inside a frame consisting of two thick vertical iron plates, 2·5 cm. by 2·5 cm. and 6 mm. thick, separated above and below the ebonite by brass distance pieces. The axis of the specimen was parallel to the steel plates, and the ebonite was fixed to one of them only, a small air gap existing between the ebonite and the other plate.

A constant current was maintained in the specimen, by balancing the p.d. across a standard 1 ohm, in series with the specimen, against a Weston cell. The p.d. between the potential leads on the specimen could then be balanced by means of a sensitive galvanometer against the p.d. in an adjustable resistance P forming part of a potentiometer circuit. The latter consisted of an accumulator in series with a fixed resistance of 20,000 ohms and the resistance P, which was made up from the following: a dial box resistance, reading to 0·1 ohm, a standard 1 ohm with shorting plug, two small home-made rheostats of thick copper wire, each of about 0·1 ohm resistance, and the necessary leads. The value of P was usually between 50 and 120 ohms, and, of course, provided a measure of the resistance of the specimen between the potential leads.

In the absence of a magnetic field, the p.d. between the potential leads on the specimen was first fairly accurately balanced against the p.d. in P by adjusting the resistances in the dial box, and then more exactly balanced by adjusting the copper wire rheostats. The magnetic field was thereupon switched on and the resulting galvanometer deflection δ was noted; its direction gave the sign of ΔR. The arrangement was calibrated by observing the galvanometer deflection D when the standard 1-ohm resistance in P was shorted, whence the value of $\Delta R/R$ was given by δ/PD. Checks on this calibration were obtained by shunting the standard 1 ohm in the main current circuit with a 50-ohm resistance and noting the corresponding galvanometer deflection D' when 50 D' should be equal to PD. At least four values of ΔR were taken with the main current in the two directions for each chosen value of H.

The field acting on the specimen when mounted for longitudinal measurement was found by means of a small semi-circular magnetic potentiometer of 542 turns, each of approximately 10 mm. × 3 mm. in area, with a distance of 1·3 cm. between the mid-points of its feet. It was joined in series with the galvanometer used in

the ΔR measurements, together with resistances necessary to permit the range of fields to be measured. It was calibrated in the usual way (Bates 1939, 1945), the necessary factors required when additional resistances were used being found by means of a mutual inductance. In the case of the transverse measurements, a small contact search-coil of 17 turns, wound in one layer on a rectangular former 2·0 cm. long and 2mm. wide, was joined in series with the galvanometer and calibrated with the aid of a mutual inductance against a standard search-coil used in known magnetic fields.

Reference has been made to the lack of precise knowledge of the actual distribution of domain vectors in a demagnetized specimen. After each individual ΔR measurement, the specimen was demagnetized by reversals and the residual field acting upon it was measured by search-coil or magnetic potentiometer. This field was caused by residual magnetism in the electromagnet and in the specimen itself. It was brought to zero by applying momentarily a neutralizing field whose value was found by trial. The first measurements of the residual field usually gave a reliable indication of the magnitude and direction of the required neutralizing field; and greater and greater field currents were switched on momentarily in the electromagnet until, finally, no residual field could be detected. Yet, however carefully a specimen was initially demagnetized from a given state of magnetic saturation, there sometimes occured extraordinary variations in the magnitude and direction of the residual field. These were on occasion so large that the specimens had to be remagnetized and demagnetized *de novo* ; accidental surges may have played a part here. Other points of interest are described under the several headings in the following section.

§3. EXPERIMENTAL RESULTS

Cobalt

The specimens were ground from a sample of No. 12 S.W.G. wire (hard-drawn) kindly supplied by Messrs. Brandhurst & Co., Ltd., The sample was 98·4% pure, the chief impurities being Ni 0·45, Fe 0·13, CaO 0·23, Mn 0·08, C 0·19, Zn 0·01, MgO 0·11, SiO_2 0·14 and S 0·02 per cent; the loss in hydrogen was 0·24 per cent. As mentioned above, the main reason for using cobalt was to test the accuracy and suitability of the method for magnetically hard materials. The results of figure 1 give $b = 4·5 \times 10^{-3}$.

Cobalt steels

Results for a 15% cobalt steel, used in the manufacture of ships' compasses and made according to the original Kelvin recipe, are given in figure 2; the material was kindly supplied by Messrs. Swift, Levick & Sons, Ltd. The material gives values of $\Delta R_{\parallel}/R_0$ and $\Delta R_{\perp}/R_0$ on opposite sides of the H axis, in accordance with theory, with a value of $b = 2·7 \times 10^{-3}$.

Two specimens, I and II, of 35% cobalt steel, kindly supplied by Messrs. Jessop & Sons, Ltd., gave the results shown in figure 3. This material has a remanent induction of 9350 gauss, a coercivity of 240 oersteds and a $(BH)_{max}$ of $0·96 \times 10^6$ gauss-oersteds. Both specimens showed very strange demagnetization characteristics. Following a set of $\Delta R_{\parallel}$ measurements, a specimen

demagnetized by reversals in the routine manner showed only weak residual magnetism. Yet on trying to reduce the latter by the momentary application of demagnetizing fields of gradually increasing intensity, the magnetization
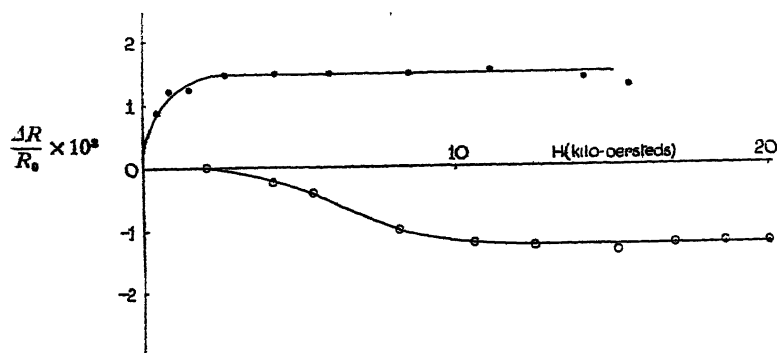
Figure 2.   Kelvin magnet steel (15% cobalt steel).

● Longitudinal field.   ☉ Transverse field.

actually increased to a maximum and then fell to zero.   Again, in the case of $\Delta R_\perp$ measurements with specimen I, sudden large deflections were often observed when the galvanometer was placed in circuit immediately after the field had been
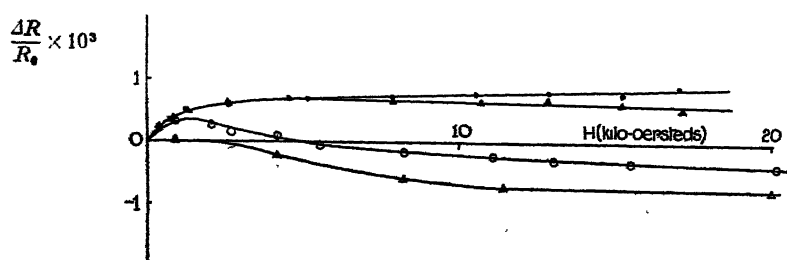
Figure 3.   35% cobalt steel.

● Longitudinal field ⎫ Specimen I.   ▲ Longitudinal field ⎫ Specimen II.
☉ Transverse field ⎭   △ Transverse field ⎭

excited.   These might be explained on the assumption that the resistance changes showed a time lag, probably connected with the strange demagnetization behaviour. The values of b for the two specimens are respectively 0.75 and 0.70×10^{-8}. It would be interesting to know whether a steel with, say, 40 per cent of cobalt give a value of b equal to zero.

Alni

The approximate composition of this alloy is iron 59, nickel 24, aluminium 13 and copper 4 per cent.   Specimens were kindly supplied by Messrs. Jessop & Sons, Ltd.   They gave a remanent induction of 6170 gauss, a coercivity of 515 oersteds and a $(BH)_{max}$ value of 1.38×10^6 gauss-oersteds.   Results of ΔR measurements are given in figure 4.   It will be noted that the graphs of $\Delta R_\parallel / R_0$ and $\Delta R'_\perp / R_0$ both lie below the axis.   The value of b is 1.15×10^{-8}.

Little significance is here attached to any difference in slope between the high-field portions of the two graphs, particularly as the specimen showed marked

directional effects with respect to the field, especially in the case of the transverse measurements, where the values of ΔR_1 for a field of given intensity (low) were about twice as great with the field in one direction as in the other. The mean values of $\Delta R_1/R_0$ are plotted in figure 4. The directional effect was much less pronounced in the longitudinal measurements, but before commencing a set of readings it was necessary to magnetize the specimen in a strong field, say 15,000 oersteds, and then demagnetize it by reversals, in order to obtain results which
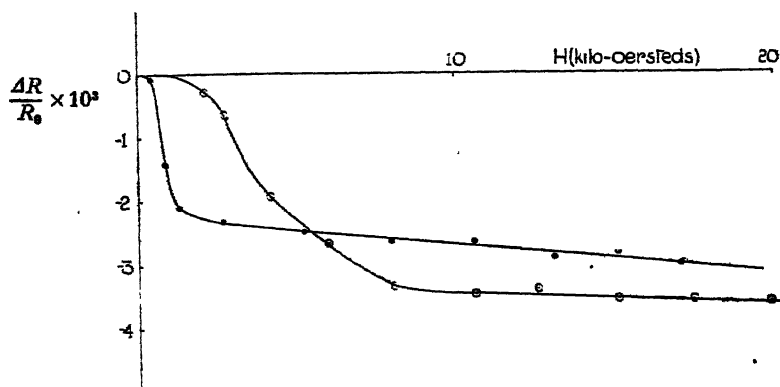
Figure 4. Nial or alni.

● Longitudinal field.   ⊙ Transverse field.

could be reproduced. This was also a feature common to steels containing much cobalt, and may perhaps be put forward in support of a suggestion by Brown (1945) that very low values of the coercivity of the single crystals of ferromagnetic metals, which have so far been investigated, have been obtained simply because these crystals have not been placed in a field, say 100,000 oersteds, sufficiently strong to break down the domains completely. Throughout the whole of the work the importance of demagnetization from high fields was constantly coming to the fore.

Alnico

The approximate constitution of alnico is iron 54, nickel 17 to 18, aluminium 11 to 10, copper 6 and cobalt 12 per cent. It has a remanent induction of 8200 gauss, a coercivity of slightly over 500 oersteds and a $(BH)_{max}$ value of 1.7×10^6 gauss-oersteds. In figure 5 are given the results for a rod of cast alnico, kindly supplied by Messrs. Jessop & Sons, Ltd. The graphs make clear the difficulty caused in some cases by lack of precise knowledge of the demagnetization field appropriate to the transverse measurements. One cannot be much in error in taking $2\pi I_s$ in figure 5 to be the applied field at the point of inflection, i.e. 5000 oersteds. This gives the value of b equal to zero, but if we take a lower value of $2\pi I_s$ we obtain a negative value of b. At high values of the longitudinal field, some uncertainty in the measurements was experienced, suggestive of the specimen being unduly strained.

Two specimens of sintered alnico, kindly supplied by Messrs. Murex Ltd., were also used. Powdered alnico cannot be sintered, of course, and the specimens were made by sintering a mixture of the constituent metals under pressure.

Sintered alnico is supposed (cf. Wulff, *Powder Metallurgy*, Cleveland, 1942) to be fine-grained and therefore stronger .than the relatively coarse-grained cast product. Since it is consolidated by inter-diffusion, it is reported to have
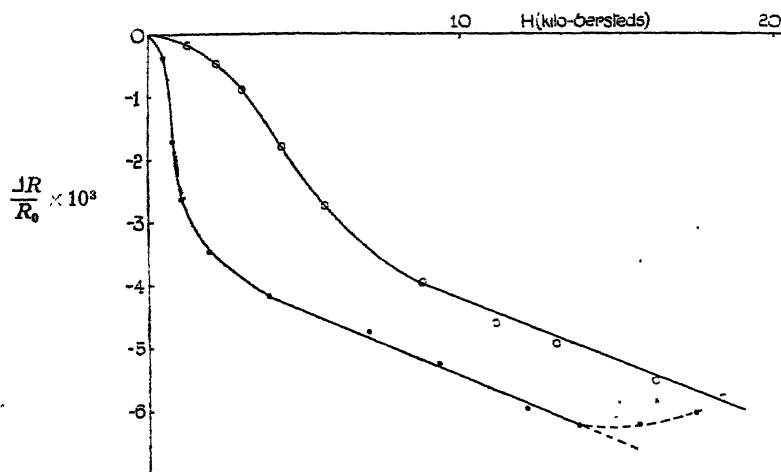
Figure 5.   Cast alnico.

● Longitudinal field.   ⊙ Transverse field.

no cracks, " cold shuts ", or segregation of impurities at grain boundaries, so that there are no consequent poles produced by physical imperfections and a more uniform flux distribution can be obtained in the material.
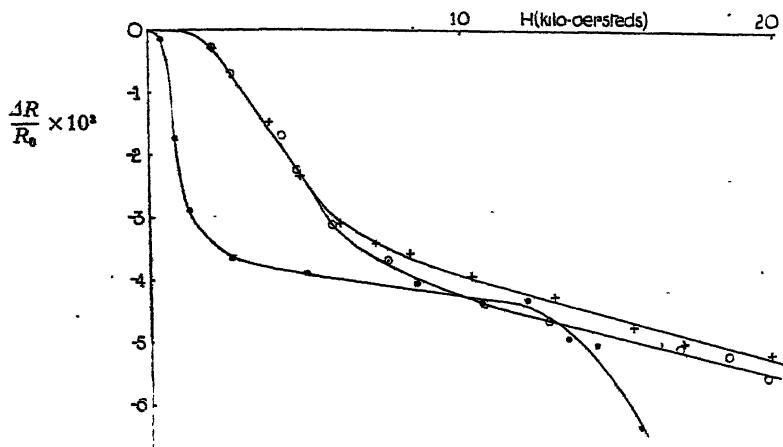
Figure 6.   Sintered alnico cut from disc.

● Longitudinal field parallel to face of disc.
⊙ Transverse field perpendicular to face of disc.
+ Transverse field perpendicular to side of disc.

The first specimen of sintered alnico, for which the results are given in figure 6, was a rod of square cross-section ground from a disc 2·1 cm. in diameter and 0·15 cm. thick, sintered under pressure perpendicular to the disc face. It was not magnetically tempered during cooling. In figure 6 the encircled points refer to·

measurements made with the applied field perpendicular to the axis of the rod and to a surface of the rod which had previously formed part of the disc surface, and the crosses refer to measurements with the field perpendicular to the axis of the rod and parallel to this surface. The two sets of measurements coincide within the limits of experimental error, so that no marked difference existed due to directional effects associated with the application of pressure during manufacture. The most striking feature of the results is, again, the smallness of the value of b; indeed, the results represented by the crosses give a small negative value for b.

The second specimen of sintered alnico was ground from a rod prepared under pressure parallel to its long axis; the results are similar to those shown in figure 6, the value of b being approximately 0.7×10^{-3}. The main difference between the results for the two specimens lies in their behaviour in the higher fields, and this is attributed to different states of strain in the two cases.

Alcomax II

The approximate constitution of this alloy is iron 50, nickel 11·5, aluminium 8·0, copper 5·5 and cobalt 25 per cent. It has a coercivity of 500 to 650 oersteds, corresponding residual induction 13,500 to 12,000 gauss and a $(BH)_{max}$ value of 4.3×10^6 gauss-oersteds, (vide *J. Sci. Instrum.* 22, 56 (1945)). Its exceptionally high coercivity, etc., are obtained as directional properties by a special heat treatment in a strong magnetic field, after casting. The above values of coercivity, etc., are measured for fields applied in the same direction as that in which the specimen was placed during this heat treatment. The magnetic properties in the favoured direction are obtained, to some extent, at the expense of those in directions perpendicular thereto, which are less good. It is reasonable *a priori* to suppose that in the unmagnetized state the large majority of the magnetic vectors are aligned either parallel or antiparallel to the favoured direction.

Resistance measurements were made with four specimens kindly supplied by Messrs. Jessop & Sons, Ltd. The results of figures 7 and 8 were obtained with specimens III and IV respectively. In the latter case measurements of $\Delta R_{\parallel}/R_0$ in low fields showed marked peculiarities, the sign and the magnitude of the deflection δ depending on the direction of the applied field; these are very clear in figure 8. It is possible that these peculiarities were due to the existence of (small) regions of exceptionally high coercivity. If one such region had its magnetic axis coincident with the specimen, the application of a longitudinal field would cause an increase or a decrease in resistance according as the magnetization of the region was decreased or increased respectively. This would account for the observed changes in δ. It should be remembered that a weak external field would give rise to a big induction inside the specimen. If the magnetic axis of the region is not coincident with that of the specimen, then we have a ready explanation of the large increase in $\Delta R/_{\perp}R_0$ in low fields as depicted in figure 8.

The suggestion finds support from the difficulties experienced in demagnetizing the specimen by reducing its residual field to zero. Usually, the necessary opposing field current to be supplied momentarily to the electromagnet could be found in a straightforward manner, because with increasing values of the demagnetizing field the potentiometer gave smaller and smaller galvanometer deflections.

which finally became zero when the correct state was reached. With specimens III and IV, however, a strange, rapid motion was superimposed upon the ordinary galvanometer deflection as the residual field became small, and it was difficult
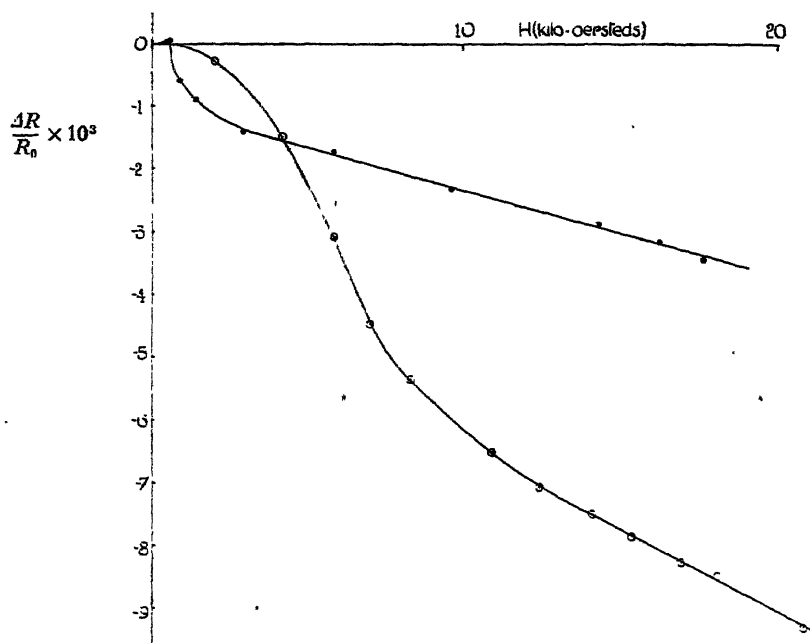
Figure 7. Alcomax II: specimen III.

● Longitudinal field.　⊙ Transverse field.
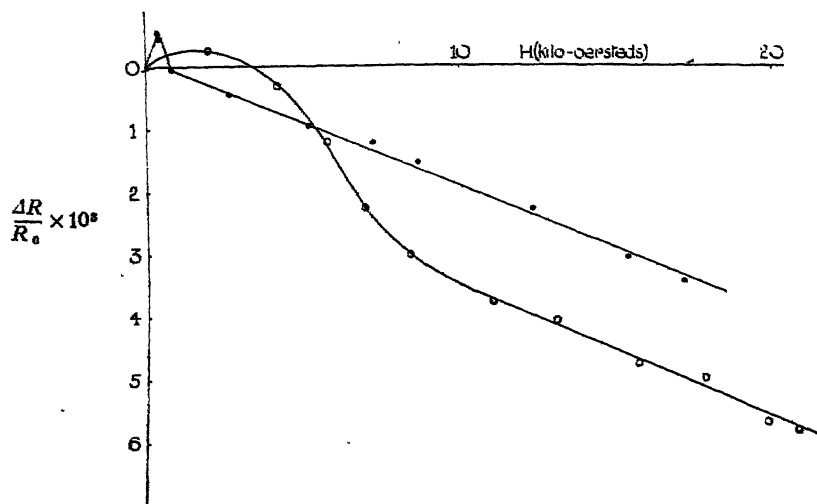
Figure 8. Alcomax II: specimen IV.

● Longitudinal field.　⊙ Transverse field.

to decide when the ordinary deflection was zero. In fact, it was often found necessary to leave a very small residual field unneutralized. The concept of a small region of high coercivity is well supported by recent experiments made in this laboratory on alcomax magnets (Bates, 1945).

The values of b from figures 7 and 8 are respectively 4·4 and 3·7 × 10⁻⁸. There are differences in the final slopes of the curves of longitudinal and transverse changes, but, in view of the uncertainty concerning the initial states of demagnetization, it is not profitable to discuss them at this stage. Resistance and corresponding magnetization and magnetostriction data for a large number of specimens of alcomax II would be necessary, for anyone with experience of high-coercivity alloys knows how wide is the variation in properties from specimen to specimen. Moreover, a special technique for measuring the slopes alone would have to be instituted to give the necessary accuracy.

It is reasonable to picture a rod formed from an ideal specimen of alcomax II as consisting of domains set with their magnetic vectors parallel to the axis of the rod, with domain sides likewise parallel to this axis. There should therefore be no abrupt change of resistance when a longitudinal field is applied, i.e., the linear portions of the upper graphs in figures 7 and 8 would, in these circumstances, pass through the origin. In ordinary specimens it is extremely unlikely that the magnetic vectors will lie exactly parallel to the geometrical axis of the rod, and that the walls of the domains will be perfectly straight and parallel. It is much more likely that there will be pockets and protuberances along the walls, which will be subject to strong lateral forces when large fields are applied. In other words, the slope of the linear portion of the $\Delta R_{\parallel}/R_0$ graph may be affected thereby, and the coarse-grained structure of alcomax is particularly liable to give rise to such phenomena. So far, no measurements have been made of magnetostriction effects, and these may play a very important part during the initial stages of magnetization.

§4. DISCUSSION

The interesting fact emerges from these experiments that the resistance of every modern high-coercivity alloy here investigated is always decreased by the application of a magnetic field, whether parallel or perpendicular to the direction of the current, apart from one or two small exceptions to this rule, when very small fields are used, which are readily understood. A similar behaviour had previously been noted only in the cases of Heusler alloy (Potter, 1932) and copper-rich alloys like constantan (Obata, 1925). Now, according to equation (7) above, the theory of domain magnetization requires that $\Delta R_{\parallel}/R_0$ and $\Delta R_{\perp}/R_0$ extrapolated to zero effective field from measurements made in strong fields should have opposite signs, which is in flat contradiction with the experimental data. An obvious suggestion to overcome this difficulty is that the magnetostriction deformations of the domain boundaries or, possibly, of the boundaries between the metal grains, play an important part. Nothing is known about the magnetostriction properties of high coercivity alloys, and it may be that their determination, now contemplated in this laboratory, will throw light on the problem.

It is also interesting that in all cases the value of b is positive or zero within the limits of experimental error; the one case, alnico, in which there are indications that b has a small negative value cannot be regarded as a proven exception to the rule. Measurements on cast and sintered alnico show no marked distinguishing features, possibly because cobalt is acting as a solvent metal in both types of material. It has been pointed out that b, which denotes the relative change in

resistance when the magnetic state of the specimen is changed from longitudinal to transverse saturation in the same effective field, should be the same for poly-crystalline as for single crystal-specimens. But we now have experimental proof that b can be zero within the limits of experimental error, and we are forced once again to the conclusion that the main resistance changes measured may be due to changes in the boundaries between the metal grains. This raises the interesting question of the reason for the marked directional properties of alcomax. Are they to be attributed to the boundaries between the grains?

Following Jones (1945), we may suppose that the boundaries in cast metals are likely to be peculiar to the alloy under consideration and likely to consist of an arrangement of atoms which is to some extent ordered. The suggestion sometimes made that there is a fundamental difference between the nature of the boundary in cast and sintered metals finds little or no support from the present experiments.

§5. ACKNOWLEDGMENTS

REFERENCES

BATES, L. F., 1939. *Modern Magnetism* (Cambridge: The University Press), p. 64; 1945. *Phil. Mag.* 36, 297.

BITTEL, H., 1937. *Ann. Phys., Lpz.*, 29, 219.

BROWN, W. F., 1945. *Rev. Mod. Phys.* 17, 15.

ENGLERT, E., 1932. *Ann. Phys., Lpz.*, 14, 589.

GERLACH, W., 1932. *Ann. Phys., Lpz.*, 12, 849.

JONES, W. D., 1945. *Principles of Powder Metallurgy* (London: Ed. Arnold & Co.), p. 74.

McKEEHAN, L. W., 1930. *Phys. Rev.* 36, 948.

OBATA, J., 1925. *Sci. Rep. Tôhoku Univ.*, Nagaoka Anniv. Vol., p. 219.

POTTER, H. H., 1931. *Proc. Roy. Soc.* A, 132, 560; 1932. *Phil. Mag.* 13, 233.

SCHARFF, GERTRUD, 1935. *Z. Phys.* 97, 73.

THIESSEN, G., 1940. *Ann. Phys., Lpz.*, 38, 167.

THE ABSORPTION OF SOUND IN A HOMO-GENEOUS POROUS MEDIUM

By R. A. SCOTT,

Manchester

ABSTRACT. Recent theories relating to the wave-propagation of acoustic disturbances in homogeneous, isotropic porous media are discussed, and expressions are given of the wave-equation and for the oscillatory pressure in the material in terms of a " velocity potential of average flow ". In their most general form these expressions are closely analogous to those which correspond to the propagation of sound in free air. The general theory is expressed in terms of two complex parameters which respectively take the place of the wave-length constant and the mean density of the air, which figure prominently in elementary acoustic theory. Expressions for the complex parameters are also given in terms of the " effective " inertia, compressibility and flow-resistance of the material.

Theory shows that numerical values of the components of the complex parameters can be obtained experimentally for practical absorbing materials at a given frequency by measurement of the attenuation constant and of the velocity of propagation of sound in the medium, together with measurement of the normal acoustic impedance at the surface of a sample of the material of effectively unlimited depth. Experimental methods suitable for these measurements are described and results of such measurements are given for a typical sample of porous material. Validity of the general theory as applied to this material is suggested by the good experimental agreement obtained on applying the theory to measurements of the normal acoustic impedance of specimens of various finite depths of the material.

From the measured values of the complex parameters referred to above, calculations of effective inertia, compressibility and porous resistance are made for the material used in the experimental work and the manner of variation of these quantities with frequency is deduced. The results throw interesting light on the mechanism of the propagation of sound in the material.

§ 1. INTRODUCTION

THE bulk of the materials used in acoustics for the absorption of sound depend for their operation on the viscous dissipation of energy in the air contained in pores in the material. From the macroscopic point of view these porous materials may be either homogeneous, as in the case of cotton-waste, felt or rock-wool, or composite, as in the case of " acoustic tiles ", which have perforated coverings of semi-rigid material. In either case the performance of the whole material depends on the manner in which acoustic disturbances are propagated in the homogeneous porous medium. In this paper a treatment is given, following the lines of those of Wintergerst, Monna, Morse, and others, of the propagation of an acoustic disturbance in three dimensions in a homogeneous, isotropic porous medium. Experiments made to check the theory are described and their results discussed in the light of the theory.

§ 2. GENERAL THEORY

Rayleigh (1929) has investigated the performance of an idealized absorber of sound, consisting of a bundle of small tubes (see also Crandall, 1927). He

extended the theories of Helmholtz and Kirchhoff that relate to the transmission of sound along tubes and was able to deduce a value for the absorption coefficient at the surface of such a material in terms of the proportion of the surface area occupied by open pores, the radius of the tubes and the physical constants of the air. This analysis has provided a qualitative basis on which to found more general theories and has led to a number of attempts to give a detailed description of the mode of action of materials such as are used in practice.

Rayleigh's theory is important in that it shows that a sound wave falling on the absorber continues, in part, to be propagated as a wave in the air in the material and that the phase velocity of the latter wave is determined by the viscosity as well as by the density and the elasticity of the air. Most porous materials used in acoustics are composed of an almost random entanglement of fine fibres in three dimensions, rather than of a collection of similar parallel tubes, and it is clear that detailed conclusions drawn from Rayleigh's theory would be inappropriate to such materials. It remains probable, however, that an acoustic disturbance is propagated as a highly attenuated wave of definite velocity through any type of homogeneous porous material, and it is demonstrated later in this paper that this surmise is correct. Wintergerst (1931), Monna (1938), and others (see Morse and Bolt, 1944) have shown that it is possible to remove the restriction of a cylindrical shape for the pore whilst still retaining wave propagation in the material. These investigators show that the attenuation and phase velocity of the wave can be expressed in terms of an effective density and elasticity of the air and of a coefficient representing the resistance offered by the material to alternating flow. In terms of these parameters, expressions can be readily deduced for the absorption coefficient and for the acoustic impedance of the material when sound falls normally on the surface. Monna has extended the theory to wave propagation in three dimensions for the case of an homogeneous, isotropic material and has deduced a wave equation to represent the disturbance. In the present paper the results of this three-dimensional type of theory are compared with the practical performance of a typical porous material (rock-wool).

It is necessary that the parameters used to express the performance of the material be clearly defined at once, since it is intended later to make deductions about the mechanism of propagation of the sound in a typical material from experimentally determined values of these parameters. For this reason the theory is briefly outlined below.

The analysis will be confined, for brevity, to the case of a medium which from the point of view of transmission of sound is homogeneous and isotropic.

The porous type of sound-absorbing material commonly consists of very many small fibres or particles which are packed together so as to leave between them interconnected air-spaces or pores of irregular shape. A sound wave falling on the bounding surface of such a material must necessarily set up disturbances, not only of the air in the pores but also of the fibres or particles themselves. The exact analysis of the performance of such a system would require a knowledge of the forces of restraint of the fibres and simultaneous solutions would then be required of the two problems:

(i) the motion of the fibres under the action of the forces exerted upon them by the moving air;

(ii) the motion of the air in the irregular pore space under the combined influences of the pressure in the initial sound wave, the movement of the fibres, and the inertial, compressive and viscous forces associated with the confined air.

When framed in this general form, the mathematics are intractable, and simplifying assumptions must be made in order to obtain a practically useful mathematical description of the progress of an acoustic disturbance in the material.

It is reasonable to suppose that the transmission of motion along the fibres by solid conduction is inappreciable, for the rigidity of a fibre is small and the motion of a portion of the fibre is localized by the restraint imposed by contact with adjacent fibres. In considering the average motion of the air in the pore space, it will be assumed that the effect of the forces due to the presence and the movements of the fibres is to modify the effective inertia of the air and to introduce into the equations of the motion of the air a dissipation term proportional to the oscillatory velocity of the air.

Let v be the instantaneous velocity of the air in the pore space averaged over an area normal to v which cuts many fibres. v may be called the average velocity of flow. Let the volume of the air in the pores be P times the total volume, where P is a factor less than unity and may be called the *geometrical porosity*. In unit area of any plane in the material, the cross-section of the pore space will tend to a constant value A by virtue of the random distribution in space of the many fibres. This quantity A is numerically equal to the porosity P, for a slice of thickness δn contains a volume $A\delta n$ of pore space, and by summing this volume through unit distance in the direction n we have $P = A\Sigma_0^1 \delta n = A$.

The equation of motion of the air in a small volume δv of the porous medium is, if we ignore second-order effects:

$$- \operatorname{grad} p . P \,\delta V = m\rho_0 P \delta V \frac{\partial}{\partial t}\left(\frac{v}{A}\right) + rvP\delta V,$$

and since $P = A$, this may be rewritten

$$- \operatorname{grad} p = m\rho_0 \frac{\partial}{\partial t}\left(\frac{v}{P}\right) + rv, \qquad \dots\dots(1)$$

in which m is the ratio of the effective density of the air in the pore space to the actual density ρ_0 of the air, p is the oscillatory pressure in the volume δV and r is a resistance coefficient analogous to the flow resistance r' in Darcy's empirical law for steady non-turbulent flow in porous media, viz.:

$$- \operatorname{grad} p = r'v. \qquad \dots\dots(2)$$

Equation (2) becomes equivalent to (1) when acceleration forces are negligible.

If we express v in terms of a velocity potential of average flow ψ, we have $v = - \operatorname{grad} \psi$, $\dot{v} = - \operatorname{grad} \dot{\psi}$. For small changes of p and therefore of ρ_0, equation (1) gives on integration

$$p = \frac{m\rho_0}{P}\dot{\psi} + r\psi. \qquad \dots\dots(3)$$

Let the condensation $\delta\rho_0/\rho_0$ be represented by s and let κ be the effective compressibility of the air in the pore space; by definition $p = \kappa s$ and (3) may therefore be rewritten

$$\kappa s = \frac{m\rho_0}{P}\dot{\psi} + r\psi. \qquad \dots\dots(4)$$

The equation of continuity is

$$P\frac{\partial \rho_0}{\partial t} + \rho_0 \operatorname{div} \mathbf{v} = 0,$$

i.e.

$$P\frac{\partial s}{\partial t} - \nabla^2 \psi = 0. \qquad \ldots\ldots(5)$$

Substituting for s in (5) from (4),

$$\nabla^2 \psi - \left(\frac{m\rho_0}{\kappa}\ddot{\psi} + \frac{rP}{\kappa}\dot{\psi}\right) = 0. \qquad \ldots\ldots(6)$$

If ψ varies periodically with pulsatance ω, (6) may be rewritten

$$\nabla^2 \psi + h^2 \psi = 0, \qquad \ldots\ldots(7)$$

where

$$h^2 = \frac{m\rho_0 \omega^2}{\kappa} - j\frac{rP\omega}{\kappa}.$$

The expression for oscillatory pressure in the pore space is given in terms of ψ by (3), which becomes for periodic motion

$$p = \left(\frac{m\rho_0}{P} - j\frac{r}{\omega}\right)\dot{\psi}$$

$$= \frac{h^2\kappa}{P\omega^2}\dot{\psi},$$

and if we write

$$\frac{h^2\kappa}{P\omega^2} \equiv \rho' \qquad \ldots\ldots(8)$$

we have

$$p = \rho'\dot{\psi}. \qquad \ldots\ldots(9)$$

Equations (7) and (9) suffice to describe any periodic acoustic disturbance in the porous medium and are analogous to the well-known equations for the propagation of unattenuated waves in a gas

$$\nabla^2 \phi + k^2 \phi = 0 \qquad \ldots\ldots(10)$$

and

$$p = \rho_0 \dot{\phi}, \qquad \ldots\ldots(11)$$

in which ϕ is the velocity potential and k is the wavelength constant.

§3. DETERMINATION OF THE MAGNITUDES OF h AND ρ' FOR A TYPICAL POROUS MEDIUM

The equations (7) and (9), which with the requisite boundary conditions define the sound field in the medium at any one frequency, involve the values of h and ρ' for the medium. These parameters are in general complex, and the magnitudes of their components depend on the working frequency. A primary concern in the experimental work described below has been to determine the magnitude of these parameters and the manner of their dependence on frequency for a typical homogeneous porous material. The material used for the tests is a variety of rock-wool known commercially by the trade name of Stillite (bulk density 0·080 gm./cm³) The material possesses good homogeneity and proves to be reasonably isotropic in its acoustic properties.

From the experimental point of view the work divides into two parts. The first part involves the determination of the propagation constant associated with the passage of a uniform plane progressive wave in the porous material. The second part involves the measurement of the acoustic impedance at the face of a substantially unlimited depth of the material. From the former measurements h is obtained and from the latter ρ'/h. The two types of measurement are quite different in nature and the description of the experimental work must therefore be divided into two parts.

Experimental determination of the propagation constant in rock-wool

The parameter h describes the spatial variation of the acoustic disturbance in the medium. Thus for a uniform, progressive, plane wave in rock-wool, the velocity potential ψ of the average motion in the pores can be written

$$\psi = \psi_0 \exp(-jhx) . \exp(j\omega t), \qquad \dots\dots(12)$$

where x represents distance in the direction of propagation of the wave. In current terminology the quantity jh is called the *propagation constant* of the wave.

If we write $jh = \alpha + j\beta$ i.e. $h = \beta - j\alpha$ and rearrange (12) we have

$$\psi = \psi_0 \exp(-\alpha x) . (j[\omega t - \beta x]),$$

and it becomes apparent that the real part of jh is the attenuation constant of the wave and the imaginary part of jh is the *wavelength constant*, $2\pi/\lambda$, of the wave in the rock-wool. The experimental determination of h therefore reduces to the determination of the attenuation and of the wavelength of a plane progressive wave in the rock-wool.

The measurements were made by setting up a progressive sound wave in a tube filled with rock-wool and exploring the sound field by means of a " probe-microphone ". The decay of intensity with distance provides a measurement of the attenuation constant α. The change of phase of the oscillatory pressure with distance provides a measurement of the wavelength, and therefore of β. The combination $\beta - j\alpha$ then constitutes the quantity h.

A diagram of the apparatus is shown in figure 1. The specimen (R) of Stillite was in the form of a prism of 5·08 cm. square cross-section and 30·5 cm. long which was mounted at one end of a square tube of metal of length 91·5 cm. Sound from a moving-coil horn-type loudspeaker unit (Vitavox type: N) (L) was led to the specimen through a rubber pipe and an initial short flared horn (P), made by moulding Plasticene into the first 5 cm. of the square tube. The remaining 55·9 cm. length of the tube was filled with an additional, similar specimen of Stillite. The absence of discontinuity in the long length of Stillite in the tube allows a sound wave from the source to pass without backward reflection, well past the measurement specimen. The length of the terminating specimen was chosen so as to ensure that a reflected sound wave from the far end of the tube is sufficiently attenuated for its amplitude at the measurement specimen to be negligible compared with that of the forward wave. The sound wave, which has a substantially plane wave-front because of the smallness of the transverse dimensions of the tube in relation to the wavelength, is therefore effectively a single progressive wave travelling away from the loudspeaker. The exploring microphone (M) consists

of a heavily encased piezo-electric microphone (rochelle-salt, with conical dia-phragm) to which sound is led through a metal "probe-tube". The electrical output from the microphone, which is proportional to the acoustic pressure at the mouth of the probe, is amplified, filtered by means of an electrical analyser (B) (Churcher and King (1934)) so as to remove such unwanted higher harmonics and any interference as may exist, and measured by means of a rectifier-type, moving-coil voltmeter (C). The voltage measuring equipment was calibrated by
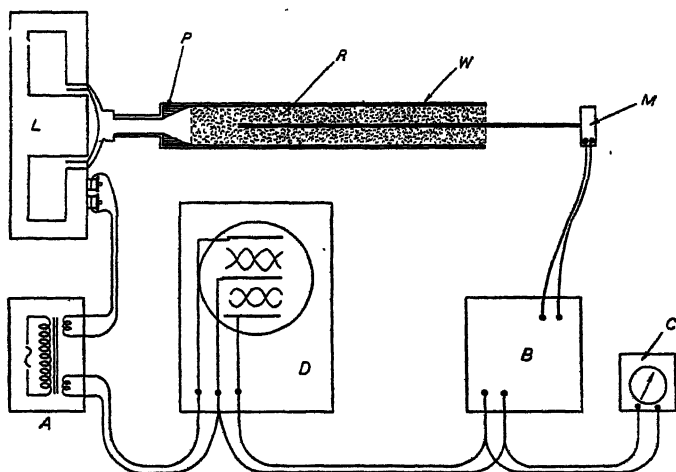
Figure 1.   Apparatus for measurement of propagation constant.

means of a known voltage supplied from the oscillator via a potentiometer to the input terminals of the amplifier.

The probe microphone was mounted on a trolley so as to allow the exploring end of the microphone tube to travel freely through the Stillite in the direction of the axis of the tube. The position of the orifice could be measured to within 1/10 mm. at the trolley by means of a reference scale. A view of the orifice end of the probe-tube is shown in figure 2; A is a bullet-shaped plug fitted into the
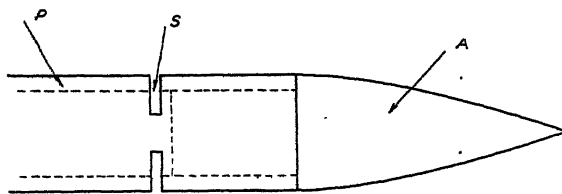
Figure 2.   Orifice of probe-tube microphone.

end of the tube and S is one of two symmetrically placed slit orifices which were made by sawing part way through the wall of the probe-tube P. The bullet-nose pushes aside the porous material as the probe enters the medium and the porous material falls back substantially to its initial position on withdrawal of the probe. In the initial stages of the work the probe was used as an open-ended tube, without the bullet-nose, and gave sets of readings which could be continually repeated by retraversing the length of the specimen. The bullet-nose was found, however,

to damage the specimen less and was therefore used in the main work. Preliminary tests with open-ended probes of different external diameters (stainless steel tubes) respectively as follows: (*a*) external diameter 0·318 cm., internal 0·214 cm. ; (*b*) external 0·241 cm., internal 0·165 cm., gave results for the attenuation constant at a given frequency which were in close agreement (e.g. within 1 per cent at 200 c/s.) and this agreement is alone sufficient reason for assuming that the presence of the exploring tube has negligible effect on the field and that the microphone pressure is a correct indication of the pressure of the wave in the porous medium.

The vertical walls of the main experimental tube (W in figure 1) are made of brass of section 5·08 cm. by 1·27 cm. and the tube is completed by removable top and bottom covers of steel 0·63 cm. thick which are clamped with intermediary thin cork gaskets to the brass walls. Suitable holes, which can be tightly closed with screws, are provided at appropriate intervals in the walls and the cover plates so as to allow a probe microphone to be used to explore the transverse sound field in the duct. When in use the tube and loudspeaker were
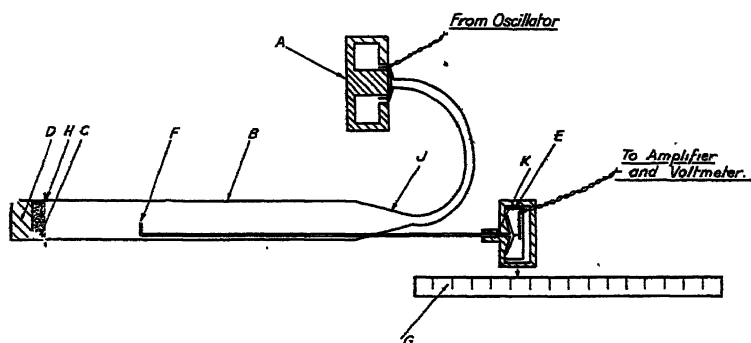
Figure 3.   Apparatus for measurement of acoustic impedance.

carefully inspected so as to ensure no appreciable leakage of sound through crevices in the apparatus either out of or into the apparatus.

Determinations of the attenuation constant were made from the slope of a graph of the pressure, on a decibel scale, plotted against distance ; thus an attenuation of Y db./cm. corresponds to an attenuation constant α given by $\alpha = 2·303Y/20$ cm.$^{-1}$ The wavelength of the sound in the porous medium was determined from the rate of change with distance of phase of the pressure at the probe mouth. The phase changes were determined with the aid of a Cossor double-beam cathode-ray oscilloscope (D) which was connected so that one pair of vertical deflection plates was operated by a voltage fed direct from the oscillator (A) and the other pair operated by the amplified voltage from the microphone. The former voltage was used to control the time base in such a fashion that each screen trace contained three half-cycles—as shown in figure 1. The voltage from the oscillator provides, of course, a reference trace of fixed phase, and this trace remains in position on the screen throughout each set of measurements. The trace due to the microphone pressure therefore moves across the screen as the microphone moves along the axis of the main tube. This particular kind of pattern was chosen because the doubling-back of the trace on itself provides in

general for each trace three sharp intersections whose lateral positions on th e
screen are quite independent of the intensity of the deflecting voltage.   Any one
of these intersections therefore provides, for each of the two voltages, a convenien t
reference point in the pattern by means of which the relative movement of on e
pattern past the other can be measured.   The interval between successiv e
intersections on the screen represents 180° of phase change of the correspondin g
voltage. , A movement of one-half wavelength of the probe orifice causes one
screen pattern to move until one intersection of the trace is replaced by an adjacen t
intersection.

The form in which the theory has been cast earlier in this paper is appropriate
only to materials which are isotropic in their acoustic properties.   Many commer-
cial porous materials, and this includes Stillite, are normally provided in sheet
form and some degree of anisotropy is introduced in the- manufacturing process.
The measurements described below of the propagation constant in Stillite were
therefore made for waves propagated in each of two directions: (a) parallel to
a line in the surface of the sheet, (b) normal to the surface of the sheet.

Results of measurements of propagation constant in rock-wool

The specimens of Stillite were prepared as follows.   Ten blocks of the material,
each in the form of a cube of 5·08 cm. side, were cut with the appropriate common
orientation from the centre of a 7·62 cm. thick sheet of Stillite.   The cubes were
weighed separately and their bulk densities determined, and of the ten, six were
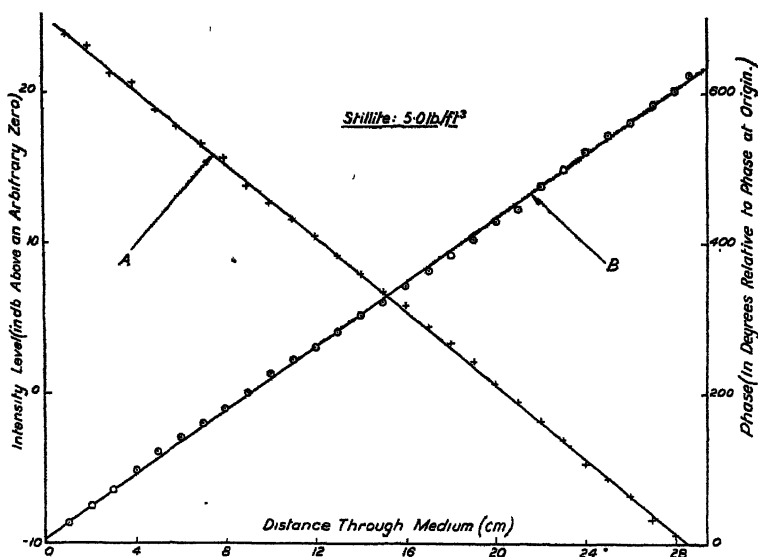
Figure 4.   Variation of intensity and phase with distance for a wave propagated in a
typical porous medium.

chosen, of closely similar density (range: 0·079 to 0·088 gm./cm³), to form the
measurement specimen.   The 5·08 cm. cubes were first arranged in the main
tube to form a specimen of type a; that is, the axis of the tube corresponded to
a direction parallel to a line in the surface of the original sheet.   Typical curves
are shown in figure 4, for a frequency of 1500 c/s., of the variation of intensity

on a decibel scale (curve A) and of phase (curve B) with distance through the sample. In each case the curve is a straight line, whence it follows that the sound is propagated with a constant phase velocity and with a constant attenuation constant. Table 1 and figure 5 show the results for the attenuation, expressed in db. per cm., and for the phase velocity for frequencies from 200 to 4000 c/s.
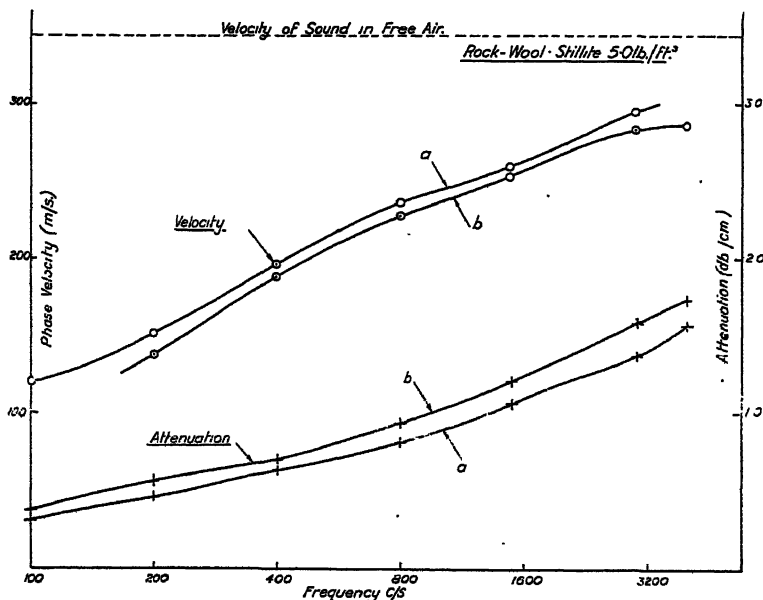
Figure 5. Attenuation and phase velocity of sound in a typical porous medium.

The values of α and of β calculated from these measurements for this direction are shown in table 2, columns 2 and 4 respectively. Each separate cube was then turned in the tube through a right angle so as to form a specimen of type **b**; that is, the axis of the tube then corresponded to a direction normal to the surface of the original sheet. The corresponding measurements of α and of β are shown

Table 1. Attenuation and phase velocity of sound waves in rock-wool

Stillite 0·080 gm./cm³

Frequency (c/s.)	Attenuation (db./cm.)			Velocity of propagation (m/s.)	
	a	**b**	Mean	**a**	**b**
100	0·301	0·368	0·344	—	120·0
200	0·452	0·546	0·499	151·6	137·9
400	0·623	0·693	0·658	196·0	188·2
800	0·800	0·937	0·868	236·5	228·0
1500	1·067	1·217	1·142	259·5	253·5
3000	1·376	1·583	1·479	295·0	283·5
4000	1·568	1·736	1·652	—	286·0

a denotes a direction parallel to surface of original sheet of material.
b denotes a direction normal to surface of original sheet of material.

in columns 3 and 5 in table 2. The combination $\alpha + j\beta$ of the above data gives the propagation constant in the appropriate direction on the Stillite. The combination $\beta - jx$ of the data gives the appropriate value of the quantity h of the theoretical analysis.

The measurements of frequency were made by comparison of the working frequency with a standard frequency of 1000·0 c/s. from a Sullivan valve-maintained tuning fork, with the aid of the Cossor double-beam cathode-ray tube.

Table 2. Attenuation constant and wavelength constant
for sound waves in rock-wool

Stillite 0·080 gm./cm³

Frequency (c/s.)	Attenuation constant (cm.⁻¹)			Wavelength constant (radians/cm.)		
	10^{-2} a	10^{-2} b	Mean	10^{-2} a	10^{-2} b	Mean
100	3·46	4·24	3·85	—	5·24	—
200	5·21	6·29	5·75	8·29	9·11	8·70
400.	7·17	7·98	7·57	12·82	13·35	13·09
800	9·21	10·78	10·00	21·25	22·05	21·65
1500	12·28	14·00	13·14	36·32	37·18	36·75
3000	15·84	18·22	17·03	63·89	66·49	65·19
4000	18·05	19·98	19·01	—	87·87	—

a denotes a direction parallel to surface of original sheet of material.
b denotes a direction normal to surface of original sheet of material.

The velocities of phase propagation in the Stillite shown in columns 4 and 5 of table 1 are obtained by dividing the pulsatance ω by β. The results show that the velocity of sound in the medium is substantially less than the velocity of sound (345 m./sec. at 22° c.) in free air and at 200 c/s. is indeed less than one-half of the free-air velocity. The measurements show some degree of acoustic anisotropy of the bulk material.

Experimental determination of the characteristic impedance in rock-wool

The characteristic acoustic impedance of a medium is the ratio of oscillatory pressure to oscillatory velocity in a progressive plane wave in the medium. Since the velocity potential for such a wave is given by an expression of the type

$$\psi = \psi_0 \exp(-jhx) \exp(j\omega t)$$

for a wave in the direction x, the expression for the characteristic impedance Z_c can be obtained with the help of (9) above. Thus

$$Z_c = p \left/ -\frac{\partial \psi}{\partial x} = -\rho' \frac{\partial \psi}{\partial t} \right/ \frac{\partial \psi}{\partial x}$$

$$= \rho' \omega / h. \qquad \qquad \ldots\ldots(13)$$

The acoustic impedance at the surface of a block of the porous medium of substantially unlimited depth is clearly identical with the characteristic impedance

of the medium: the experimental methods available for the measurement of acoustic impedance therefore lead to the determination of ρ'/h and thus to ρ'.

The apparatus used for this measurement of impedance has been described elsewhere (Scott, 1946). It is a form of the apparatus devised in 1913 by H.O. Taylor for the measurement of absorption coefficients of acoustic materials and developed by Paris for the measurement of acoustic impedance. The method involves the investigation of the stationary-wave pattern set up in a tube, driven by a source of sound at one end and terminated by the flat surface of the sample at the other end. The sound field along the axis of the tube is explored by means of a probe microphone which receives sound from the field through a tube of small diameter. The ratio of maximum to minimum pressure and of the position of the first minimum relative to the face of the sample are determined and the resistive (R) and reactive (X) components of the acoustic impedance are obtained from the results.

The apparatus has been designed in particular to provide sound waves of high uniformity in the tube and to allow precise measurements, i.e. to within 0·1 mm., of the position of the minima. In interpreting the results, the effect of attenuation in the tube in modifying the positions and magnitudes of the minima has been considered and corrections applied where necessary. Experience with the apparatus has shown it to be convenient and reliable in use. Measurements made on different occasions on a particular sample show a consistency of 1 to 2 per cent for both resistive and reactive components. Specimens for use in the apparatus must be cut in the form of cylinders 4·45 cm. in diameter so as to fit the tube. The material was cut carefully with a flat terminating surface normal to the axis of the cylinder and was positioned carefully in the tube.

A diagram of the apparatus is shown in figure 3. A is the source of sound, B the tube in which the stationary wave is formed, C is the sample of material backed by a heavy brass piston D, E is the microphone and F the orifice of the probe-tube. The position of the orifice of the probe-tube is determined by means of the scale G located near the microphone.

For the measurements of the characteristic impedance of rock-wool, the six specially chosen cubes of Stillite used in the measurement of propagation constant were divided into two sets of three. From the specimens of the first set, cylinders 4·45 cm. in diameter and 5·08 cm. long were cut, with the axes of the cylinders along the direction a referred to in the earlier experimental work. From the specimens of the other set, similar cylinders were cut with axes along the direction b. The three specimens with the a axis were then fitted in one end of the cylindrical specimen tube (27·4 cm. long) of the impedance-measuring apparatus so as to present a total depth of material of 15·24 cm. The results of the measurements of α already quoted, show that even at 200 c/s. a wave passing into the specimen and reflected completely at the far end would be diminished in amplitude to one-sixth on rearrival at the beginning of the specimen. The 15·24 cm. depth of Stillite was backed by a long (25 cm.) cylinder of similar material cut from the original block, though not as carefully selected; this continuation therefore largely suppresses any tendency for a reflected wave in the specimen to affect the sound field in the measurement tube of the impedance-measuring apparatus. This 15·24 cm. length of Stillite with the Stillite backing

therefore behaves to a sound wave incident on an end surface, substantially as an infinitely deep specimen. The real and imaginary components of the acoustic impedance at the surface of the composite specimen were measured at frequencies in the range 200 to 4000 c/s. and the results are shown in columns 2 and 5 in table 3. The three component specimens which had been cut with axes in the **a** direction were replaced by the cylinders cut with axes in the **b** direction. The real and imaginary components of the impedance were measured for this sample and the results are shown in columns 3 and 6 of table 3.

Table 3. Characteristic acoustic impedance for sound waves in rock-wool

Stillite 0·080 gm./cm³

Frequency (c/s.)	Resistive component R			Reactive component X		
	a	**b**	Mean	**a**	**b**	Mean
100	95·9	97·1	96·5	77·9	83·9	80·9
200	74·9	71·3	73·1	51·1	50·7	50·9
400	62·2	61·7	62·0	32·1	31·6	31·8
800	59·2	53·8	56·5	19·4	18·5	19·0
1500	52·8	49·5	51·2	12·5	11·3	11·9
3000	50·8	48·3	49·6	5·4	4·1	4·7
4000	51·2	48·2	49·7	4·8	5·0	4·9
5000	49·2	47·6	48·4	2·0	1·4	1·7

a denotes direction of propagation parallel to surface of sheet.
b denotes direction of propagation normal to surface of sheet.

Comparison of the measured values of acoustic impedance of the bulk material for the two directions **a** and **b** in the original sheet again shows acoustic anisotropy of the Stillite. Since the anisotropy is small, both in the measurements of the propagation constant and of the characteristic impedance, mean values have been used in the comparison, which follows, of the experimental results with the theoretical analysis.

§4. GENERAL DISCUSSION OF EXPERIMENTAL RESULTS

The measurements of attenuation and of velocity of propagation of sound in the Stillite allow us to calculate h at the appropriate working frequency, and the measurements of the characteristic impedance of the Stillite allow us to deduce the corresponding value of ρ'. The values of the parameters h and ρ', determined in this manner, are shown in table 4; these values are calculated from the mean of the measurements in directions **a** and **b** of α and of β. It should be remembered that for the propagation of sound in free air, h becomes the wavelength constant usually denoted by k ($= 2\pi/$wavelength in free-air), and that ρ' becomes ρ_0 ($= 0·001205$ gm./cm³, for these measurements). These quantities have been entered in table 4, columns 6 and 7, for purposes of comparison with h and ρ'.

In the work described above, it has been assumed that acoustic disturbances in porous materials can be represented completely by two parameters h and ρ',

and it has then been shown experimentally, for the particular case of a variety of rock-wool, how measurements of these quantities may be obtained for frequencies in the range 200 to 4000 c/s. If the initial assumptions are true, it should now be possible to calculate, for example, the acoustic impedance Z_l of a layer of known thickness (l) of this material—Stillite—when backed by a substantially rigid backing. Thus it is easy to show (Scott, 1946 a) with these assumptions that

$$Z_l = Z_c \coth (jhl) = \frac{\rho' \omega}{h} \coth (jhl).$$

Using the values of ρ' and h of table 4, Z_l has been calculated for thicknesses of 2·54, 5·08 and 10·16 cm. from this formula, and the calculated values are shown in figure 6, plotted against frequency. Corresponding measurements of Z_l were made with the impedance-measuring apparatus, on the samples used earlier for the determination of ρ' and h, and these are plotted in figure 6.

At frequencies of 400 c/s. and above, the agreement between measured and deduced values of Z_l is excellent. At 200 c/s. the agreement, particularly

Table 4. Experimental values of parameters h and ρ' for rock-wool

Stillite 0·080 gm./cm.³

Frequency (c/s.)	h		ρ'		k	ρ_0
	Real	Imag.	Real	Imag.		
200	0·0870	−0·0575	0·002734	0·006870	0·03642	0·001205
400	0·1309	−0·0757	0·002669	0·003527	0·07284	0·001205
800	0·2165	−0·0100	0·002057	0·001941	0·1457	0·001205
1500	0·3675	−0·1314	0·001831	0·001178	0·2732	0·001205
3000	0·6519	−0·1703	0·001672	0·000611	0·5463	0·001205
4000	0·8504	−0·1901	0·001645	0·000542	0·7284	0·001205

in regard to the resistance, is not so good. It is probable that this discrepancy is due to experimental error in the measurement of h, since accuracy of measurement of the wavelength constant β was limited by the shortness, in terms of wavelength, of the specimen used.

The magnitudes of the attenuation and of the velocity of propagation of sound in the rock-wool are of interest in themselves to all who use such acoustic materials. Porous materials are most commonly used in acoustics for the production of wall linings which possess a low coefficient of reflection, and this property of the lining depends at least as much upon the characteristic impedance of the medium, which together with that of the air determines the amplitude of the wave immediately reflected from the lining surface, as upon the capacity of the medium itself to absorb energy. A secondary, and at the moment much less common, use of porous materials is for the purpose of sound insulation. For this latter purpose, the important factor is the ratio of the sound pressures on two sides of a given layer of the material and this is primarily dependent on the attenuation of sound energy in the material. The measurements of attenuation recorded in table 1 show rates of attenuation which, when converted to the appropriate units, correspond to 30 db. per foot thickness of the layer at 1000 c/s.

Rock-wool of the density (0·080 gm./cm³) used in these tests is not particularly useful as an insulating partition, except at high frequencies in the acoustic range. Similar tests which have been made during the present investigation of more tightly compressed material of the same type, show much higher attenuation, e.g. for Stillite of density 0·384 gm./cm.³ the attenuation is 75 db./ft. at 200 c/s. and 120 db./ft. at 1000 c/s.

The measurements of the velocity of propagation of sound in the porous material are particularly interesting. At frequencies of 100 c/s. the velocity is about one-third of that for free air (345 m/s. under corresponding conditions), whereas at frequencies in the region above 4000 c/s., the velocity in the Stillite
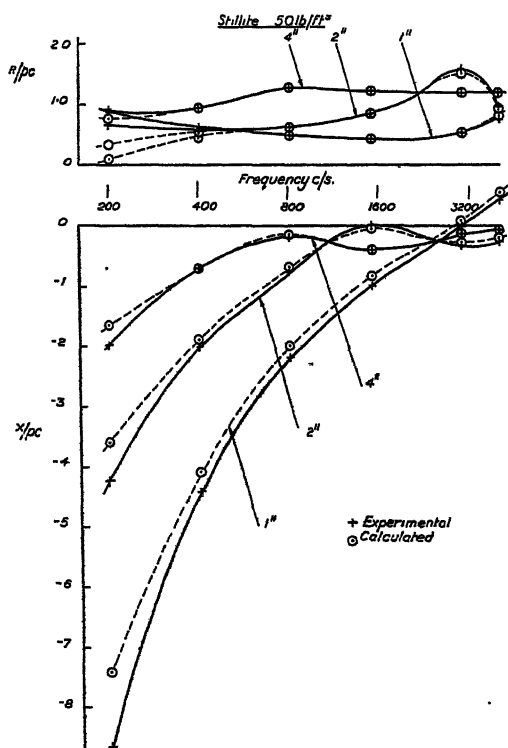
Figure 6.   Comparison of measured and calculated values of acoustic impedance of rock-wool samples of various thicknesses.

approaches that in free air. This considerable dispersion of velocity with frequency is closely analogous to anomalous dispersion of light and is, indeed, commonly found in cases of wave-propagation in dissipative media. It is of interest to compare the directly measured values of the velocity of sound in this sample of rock-wool with the rough estimates of the velocity of propagation of sound in felt as made by Crandall many years ago from an analysis of W. C. Sabine's measurements of the absorption coefficient of layers of wool-felt (Crandall, 1927). Crandall obtained values ranging from 91 m/s. at 480 c/s. to 127 m/s. at 2048 c/s. These results show the same upward trend of velocity with rising frequency as has been measured in the present work.

§ 5. INTERPRETATION OF THE MEASURED VALUES OF h AND OF ρ' IN TERMS OF THE GENERAL THEORY

In the theory of Monna, Wintergerst and others which has been outlined earlier in this paper, it has been shown that with certain simplifying assumptions it is possible to construct a wave equation for the propagation of sound in a homogeneous, isotropic porous material in terms of parameters P, m, κ and r which represent respectively the geometrical porosity factor, the effective mass of the contained air, the effective compressibility of the contained air and the porous resistance for alternating flow of the material.

The geometrical porosity factor P can be determined by direct methods. For example, if ρ_B is the bulk density of the material, measured in terms of the weight of a specimen of known volume, and if ρ_s is the density of the material

Table 5. Calculated values of parameters m, r,
$|\kappa|$, amp κ and $|\gamma'|$ for rock-wool

Stillite 0·080 gm./cm³

Frequency (c/s.)	m	r (gm.cm.$^{-3}$ sec.$^{-1}$)	$|\kappa|$	amp. κ (degrees)	$|\gamma'|$
200	2·20	8·63	1·041	−1·4	1·03
400	1·83	8·86	1·131	+2·8	1·12
800	1·66	9·76	1·218	+6·3	1·20
1500	1·47	11·06	1·232	−6·6	1·22
3000	1·35	11·52	1·352	+9·2	1·33
4000	1·32	13·62	1·397	−7·0	1·38

of the solid fibres determined on a crushed specimen with the aid of a specific-gravity bottle, the porosity P which is defined as the proportion of the voids to the total volume is given by

$$P=(\rho_s-\rho_B)/\rho_s=1-\rho_B/\rho_s.$$

The mean bulk density of the samples of Stillite used in the experimental work is 0·080 gm./cm³ and the density of the fibres 2·65 gm./cm³, whence $P=0·970$.

The experimental determinations of h and of ρ' for the porous material, allow the calculation of the values of the parameters m, κ and r. Thus

$$h^2=\frac{m\omega^2\rho_0}{\kappa}-j\frac{\omega r P}{\kappa},$$

$$\rho'=\frac{\kappa}{P\omega^2}\left(\frac{m\omega^2\rho_0}{\kappa}-j\frac{\omega r P}{\kappa}\right)=\frac{m\rho_0}{P}-j\frac{r}{\omega}.$$

Hence

$$\kappa=P\omega^2\rho'/h^2,$$

$$m=\text{Real part of } P\rho'/\rho_0,$$

$$r=\text{Imaginary part of } \rho'\omega.$$

These quantities have been calculated from the data of table 4 and are shown plotted against frequency in figure 7. None of these quantities remains constant as the frequency is varied. At low frequencies (200 c/s.) the effective mass

of the air in the pores of the material is roughly double that of a similar quantity of unrestricted air. At higher frequencies (3000 c/s.) the effective mass decreases and, indeed, approaches the value appropriate to unrestricted air. The value of r, the effective porous resistance coefficient of the material, increases with frequency. At 200 c/s. r has a value (8·63 gm. cm.$^{-3}$ sec.$^{-1}$) which is closely the same as the steady flow porous resistance coefficient (8·6 gm. cm.$^{-3}$ sec.$^{-1}$) of Darcy's law. The latter coefficient was determined by measuring the pressure differences across samples used in the earlier tests, for known rates of flow of air through the sample. The velocity of flow was never allowed to exceed 1·1 cm./sec. which is a velocity of about the same order as the oscillatory velocities in the sound wave during the acoustic measurements. The relation between pressure difference across sample and velocity of flow through sample was shown to be
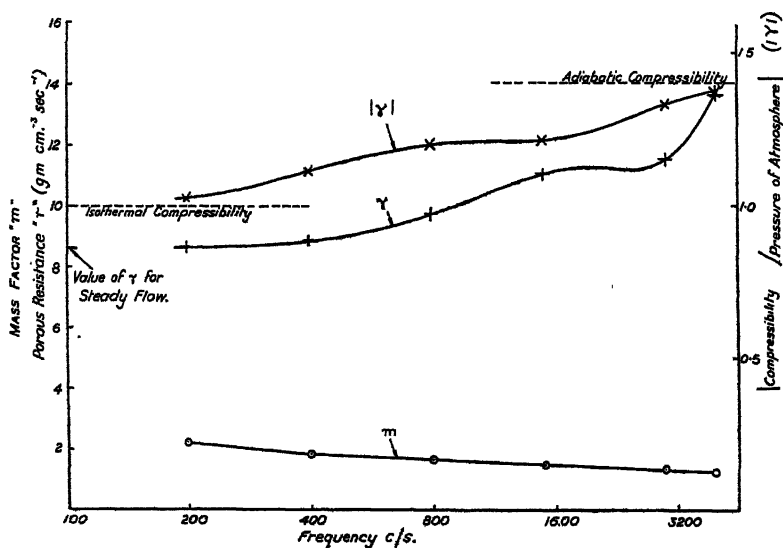
Figure 7. Variation wtih frequency of parameters m, r and $|r|$.

linear for the velocities used and the value quoted above is the mean of measurements (8·0 and 9·2 respectively) for flow parallel to and normal to the original surface of the Stillite sheet.

The compressibility of the air in a sound wave propagated in air is ordinarily the adiabatic elasticity $\gamma\Pi$, in which γ is the ratio of the specific heats of the air (1·403) and Π is the atmospheric pressure. For a condensation which takes place under isothermal conditions, the elasticity of the air is numerically equal to the atmospheric pressure Π. The effective compressibility κ determined from h and ρ' for the sample of rock-wool is in general a complex quantity, but at all frequencies the phase angle is small. The values of $|\gamma'|(=|\kappa'|/\Pi)$ are plotted against frequency in figure 7. At the low-frequency end of the range, the compressibility $|\kappa|$ tends to Π and at the high-frequency end it tends to 1·4 Π.

The construction of a theory relating to the propagation of an acoustic disturbance in terms of effective values of inertia, compressibility and resistance, involves only a partial view of the behaviour of the air in the pores. The effective

values of these various quantities depend for their magnitude on the detailed movements of the air and of the fibres. Reference has already been made to the difficulties of an exact mathematical analysis of the individual motions; nevertheless, qualitative deductions are possible about the values to be expected for the parameters.

The effective values of inertia, resistance (and, to some extent, compressibility), will depend on the distribution of velocity across the air space between the fibres. At the surface of a fibre there will be no relative motion between the air and the fibre. For positions well away from the fibre surface, the velocity will tend to a uniform value. The velocity over the intervening sections will depend on the frequency, on the kinematic viscosity of the air, and on the geometrical arrangement of the fibres (vide: the classical analyses by Helmholtz and Kirchhoff of the propagation of waves in tubes, given in Rayleigh's *Theory of Sound*, vol. 2). At high frequency the velocity gradient will tend to be concentrated near the surface of the fibres, whereas at lower frequencies this phenomenon will be less pronounced and the flow become much as in the case of steady flow.

At frequencies high in the acoustic range, the bulk of the air in the pores will tend to move as a core in which the motions are all in common phase. It may therefore be expected that the effective inertia will depend on the mass of the bulk of the air present in the pores and correspond with the normal density of the air (i.e. $m = 1$). At frequencies lower in the range, the proportion of the total mass of air that moves with phase different from that of the core will become larger and the inertia due to the motion of the air alone will correspond in consequence to an effective density of air less than the true density.

In the considerations advanced so far, the fibres have been assumed stationary. The viscous and accelerative forces which are imposed on the fibres by the motion of the air lead to movements of the fibres themselves, and these movements imply an increase in the effective inertia of the air. At frequencies of the order of 4000 c/s. it appears to be legitimate to assume that the inertia of the fibres is so high that their motion may be neglected; at frequencies of 200 c/s., the fibres take up an appreciable proportion of the velocity of the air as has been shown by an auxiliary experiment. Thus observations have been made of (*a*) the amplitude of fibres and (*b*) the amplitude of the air as traced out by MgO smoke particles, for a wave set up in a small wad of rock-wool of the type used in the earlier experiments. The results show that at 200 to 400 c/s. motions of the fibres are of the order of 1% of the mean motion of the air. In view of the high ratio of the mass of the fibres to the mass of the air (about 60 : 1 in the present case) the comparatively small motion of the fibres can contribute to the total inertia an amount about equal to that due to the air alone. This increase in the effective inertia more than outweighs the decrease due to lack of uniformity in phase of the motion of the air and accounts for the high experimental values of the corresponding parameter m (approximately 2 at 200 c/s.).

The tendency of the gradient of velocity to steepen at the fibres as the frequency is increased leads to a viscous force on the fibres, and therefore to a porous resistance coefficient, which increases with frequency. Apart from such dissipation as may occur due to the motion of the fibres themselves, the observed

rise of porous resistance with frequency can be explained by the change of velocity distribution of the air that accompanies increase of frequency.

The effective compressibility of the air depends jointly on the elasticity of the air itself and upon the bulk elasticity of the fibres. Since the latter elasticity must be very high, the effective compressibility may be considered as that of the air alone. The presence of the fibres will, however, affect the temperature of the air during the cycle of compression and rarefaction and it is therefore necessary to consider whether the compressibility corresponds to isothermal or to adiabatic conditions. The heat capacity of the fibres is approximately fifty times that of the included air and hence the fibres form a heat reservoir easily sufficient to retain the air at substantially constant temperature. The thermal conditions during the cycle therefore depend upon whether the air is in sufficiently intimate contact with the surrounding fibres for equalization of temperature to be complete in a small fraction of a cycle. The experimental values found for κ may therefore be interpreted as indicating that at 200 c/s. equalization is almost complete and the compression tends to be isothermal, whilst at high frequencies (4000 c/s.) the cycle occupies such a short time that the compression tends to be adiabatic. At frequencies intermediate between the two extremes it is to be expected that the maximum pressure of the cycle will be in advance of the phase of the condensation, on account of the thermal loss to the walls. The experimental values of κ show just such a lead to middle and high frequencies, but show a lag (which is physically impossible from energy considerations) at the lowest frequency used (200 c/s.). It has already been indicated in another context that the measurements of phase velocity are a little in doubt at this frequency and it is probable that such an experimental error is responsible for the discrepant sign of amp κ.

§6. CONCLUSIONS

A theoretical analysis based on the assumption that the effect of the fibres can be regarded as modifying the inertia and compressibility of the air and as introducing a dissipative force into the equation of motion, shows that an acoustic disturbance of fixed frequency is propagated in a porous medium as an attenuated wave. Explorations of the sound field in a typical porous material (rock-wool) have shown that such wave-propagation is characteristic of the material, and have led to direct determination of the propagation constant of the waves in this material. The measured values of the propagation constant, together with those of the characteristic impedance, have been used to calculate the acoustic impedance of layers of the material of thickness 2·54 cm., 5·08 cm. and 10·16 cm. when backed by a substantially rigid wall. The calculated values of the impedance agree sufficiently well with the corresponding measured values to leave little doubt that description of the performance of sound in the material in terms of a simple, attenuated sound wave is adequate.

The experimentally determined values of the propagation constant, and of the characteristic impedance have been used further to find the effective magnitudes of the inertia, compressibility and porous resistance, characteristic of the sample of rock-wool. The results show that each of these quantities changes with frequency and the reason for this variation is discussed. Thus the effective

inertia of the air at low frequencies corresponds to a density of the air of about twice its ordinary value, and the increase is attributable to movements of the fibres; at higher frequencies such movement is small, and the effective inertia of the air has a lower value. The compressibility of the air, that is, the ratio of the oscillatory pressure to the condensation, proves in general to be a complex quantity with a small phase angle. The reason for the phase factor is discussed. The modulus of the compressibility has a value which at low frequencies corresponds closely to the isothermal value and at high frequencies to the adiabatic value. The porous resistance factor associated with sound waves in the material is shown to approximate in value to the steady-flow resistance for frequencies of about 200 c/s., but thereafter to rise steadily with frequency.

The above conclusions about the variation with frequency of the parameters which are associated with sound propagation in rock-wool, qualify the conclusions arrived at by Morse from an interpretation of the results of the impedance measurements of Beranek: namely, that values of m, κ and r that are independent of frequency lead to a satisfactory description of the sound field in a porous material.*

ACKNOWLEDGMENTS

REFERENCES

CHURCHER and KING, 1934. *J. Inst. Elect. Engrs.* 75, 411.
CRANDALL, 1927. *Vibrating Systems and Sound* (London: Macmillan), chapters 3 and 5, and p. 197.
MONNA, 1938. *Physica*, 5, 129.
MORSE and BOLT, 1944 a. *Rev. Mod. Phys.* 16, 89 ; 1944 b. *Ibid.* 99.
RAYLEIGH, Lord, 1929. *Theory of Sound* (London: Macmillan), vol. ii, 328.
TAYLOR, 1913. *Phys. Rev.* 2, 170.
WINTERGERST, 1931. *Schalltechnik*, 4, 85.

* In a recent publication by Morse and Bolt (1944 b) the author indicates that some variation of the parameters with frequency must be postulated for certain materials.

THE APPLICATION OF ULTRA-SHORT-WAVE DIRECTION FINDING TO RADIO SOUNDING BALLOONS

By R. L. SMITH-ROSE and H. G. HOPKINS,

Radio Division, National Physical Laboratory

ABSTRACT. The paper describes the early portion of an investigation into the possibilities of determining wind velocities at high levels by radio methods. It was pre-, pared in the spring of 1939, but publication has been delayed owing to the important part which this technique has contributed to meteorology during the war. In the intervening six years considerable progress has been made in the application of the basic principles described in this paper. An account of this later work, leading to the practical techniques now used by the Meteorological Office, will be published later.

The position of a radio-sounding balloon transmitter can be determined by the aid of two direction-finders set at a suitable distance apart, together with a measurement of the altitude, which can be obtained from the barometric pressure and the thermometer units attached to the apparatus. The present paper is concerned with an investigation of the accuracy of the direction-finding technique used for the observation of bearings on the signals from radio sounding balloons during actual flight. From a knowledge of the precision of such radio bearings, the accuracy of position location of the balloon can be computed with the aid of curves given in the paper.

Preliminary experiments conducted with direction-finders of both the closed-loop and spaced-aerial types on wave-lengths between 8 and 10 metres (frequencies 30 to 38 Mc/s.) showed that a reasonably high instrumental accuracy was obtainable ; but that in order to utilize this to the best advantage it was necessary to install the direction-finders on very satisfactory sites. Accordingly, the experiments were continued with the aid of specially designed direction-finders of the rotating-loop type installed on two superior sites situated at the ends of a suitable base-line for balloon location work. As a result of the calibration of these direction-finding installations, it was found that the observational accuracy was of the order of $0°·1$. The maximum error in bearing obtained in the course of the ground calibrations at a short distance was about $1°$, but successive calibrations of the installations over a three-monthly period disclosed a variation in this site error of about $0°·5$.

The results of tests carried out on some twenty balloon flights show that, by taking mean observed bearings over each minute period, 95 % of the bearings are correct to within $2°$ and nearly 80 % correct to within $1°$; the probable error of such mean observed bearings, after correction for the ground calibration, is about $0°·3$.

Foreword

THIS paper was originally prepared in the spring of 1939, but publication has been delayed owing to the important part which this technique has contributed to meteorology during the war.

§1. INTRODUCTION

During several years prior to 1939, considerable progress had been made in the use of free sounding balloons with small radio transmitters attached for exploring the meteorological conditions of the atmosphere. The transmitters which formed the prototype of those used at the present time were designed to

emit radio waves, with suitable superimposed characteristics dependent upon the pressure, temperature and humidity of the air in which the balloon is situated. At the ground receiving station, measurements made on the incoming signals led to a direct interpretation of these three meteorological quantities. A description of the type of transmitter developed in this country has been given elsewhere by H. A. Thomas (1938), and preliminary experiments have shown this to be suitable for exploring aerological conditions up to altitudes of some 45,000 feet. These experiments have indicated that the necessary radio transmissions can be carried out very effectively on a wave-length in the region of 10 metres (frequency 30 Mc/s.). While much of the early experimental work was carried out mainly on a wave-length of 8·5 metres (34 Mc/s.), the standardized type of balloon transmitter used in recent years by the Meteorological Office operates within the international wave-band for radio sonde work, viz., 10·7 to 10·9 metres (27·5 to 28·0 Mc/s.).

In the past, a further aerological condition—the speed and direction of the wind at various heights—had been determined by making visual observations on free sounding balloons with theodolites. The attachment of a radio transmitter to the balloon opened up the possibility of using a radio method of determining the wind velocity, this method having the great advantage of not being limited by conditions of visibility. This limitation implies that the theodolite method is useless for determining wind conditions above the prevailing cloud level. With a view to exploring the possibilities of such a method, an investigation was conducted during 1938–39 on the accuracy attainable in the use of radio direction-finders on the ultra-short-wave signals emitted by the balloon transmitters. A preliminary investigation of radio direction finding on wave-lengths between 6 and 10 metres (frequencies 30 to 50 Mc/s.) using signals emitted from ground transmitters has been published previously (Smith-Rose and Hopkins, 1938), and the present paper deals with the extension of this work to observations made on signals from balloon transmitters.

§2. THE LOCATION OF A FREE BALLOON EQUIPPED WITH A RADIO TRANSMITTER

There are two straightforward methods of determining the position of a radio transmitter, given its height above the earth's surface. The first method is to measure the azimuth and angle of elevation of the balloon transmitter by observations taken at a single ground station. In the second method two direction-finders located at a known distance apart are employed to determine the position of the projection of the balloon on the ground.

(a) *Method involving the measurement of elevation*

The first method has the obvious advantage of requiring only a single observing station; but, while the measurement of azimuth can be carried out with the existing pattern of radio direction-finder, the determination of the angle of elevation is much more difficult. This matter has been previously referred to briefly by H. Diamond, W. S. Hinman and F. W. Dunmore (1937), but no experiments on the subject appear so far to have been carried out. In a British Patent, H. A. Thomas (1937) has described one method of determining azimuth and elevation with a single instrument. This is of the Adcock direction-finder type, with a pair

of vertical aerials spaced more than one wave-length apart.   During the rotation of such an aerial system about a central vertical axis, the received signal passes through a series of minimum intensity positions, two of which indicate the azimuth and the others indicate the elevation of the transmitter.   Practical considerations limit the application of such an instrument to wave-lengths of about 3 metres or below (frequencies of 100 Mc/s. or above), and the technique of the subject on these wave-lengths is accompanied by many difficulties and inaccuracies, the removal of which will require considerable investigation and development.   It was therefore considered to be more expedient to devote attention in the first instance to the two-direction-finder method, since this made use of instruments and a technique which were already in an advanced stage of development and which were therefore expected to produce more rapid results.

(b) *Method involving two direction-finders* .

At the time of the commencement of this investigation (1938) it was known that radio direction-finders were subject to various classes of error, some of which are directly associated with the instrument.   As a result of various phenomena accompanying the propagation of electric waves either along the ground or through the atmosphere, the error of bearings observed on distant transmitting stations may at times be considerably greater than the instrumental error, and this frequently places a limit on the accuracy obtainable in the practical use of direction-finders.   A considerable amount of modern research in direction finding has been devoted to the removal of errors resulting from the incidence of the waves at an appreciable angle of elevation above the ground.   This is precisely the case in which direction-finders are employed to locate a transmitter attached to a free balloon, and it is a matter of experiment to ascertain exactly what accuracy in bearing is attainable under these conditions.   From a knowledge of the accuracy in bearing obtainable with a single direction-finder, the consequent accuracy of position determination can be calculated.

When simultaneous bearings are taken on a fixed transmitter with two direction-finders having a known liability to error, the position of the transmitter lies within a quadrilateral which defines the maximum error in position under the conditions of the observations.   The size and shape of this quadrilateral vary with the distance and direction of the transmitter from the direction-finders.   The accuracy of position finding by this means has been studied by Corriez and Perlot with the same objective as that of the present investigation, and their results were given in a paper published in 1935.   Based on the analysis outlined in this paper, the authors have calculated the curves drawn in figures 1 *a* and 1 *b*, which show the error in location of the transmitter as a function of its position for maximum bearing errors of $\pm 0°\cdot 5$ and 1° respectively.   For clarity, each diagram is drawn in two halves with separate scales of distance.   These graphs are curves of equal error expressed as the ratio of the error in position to the length of the base line between the two D.F. stations.   It will be seen that the curve of optimum accuracy is a circle of diameter equal to the base-line.   Thus when the balloon transmitter is on the perpendicular bisector of the base-line, the maximum accuracy is obtained at a distance from this line equal to half its length.   Under these conditions the maximum possible error in position measurement is 0·9% and 1·8% of the base-

line for the two D.F. bearing accuracies of $\pm 0°{\cdot}5$, and $1°{\cdot}0$. In the first of these cases, the error in position increases to 7·5 and 28·5% for distances along the perpendicular bisector of twice and four times the base-line respectively. If the balloon transmitter remains at a constant distance from the centre of the base-line, but changes in azimuth, then the maximum error in position is a function of this position, as is shown by the graphs. On the other hand, if the path of the balloon should coincide with one of the curves in the figure, then the possible error in its
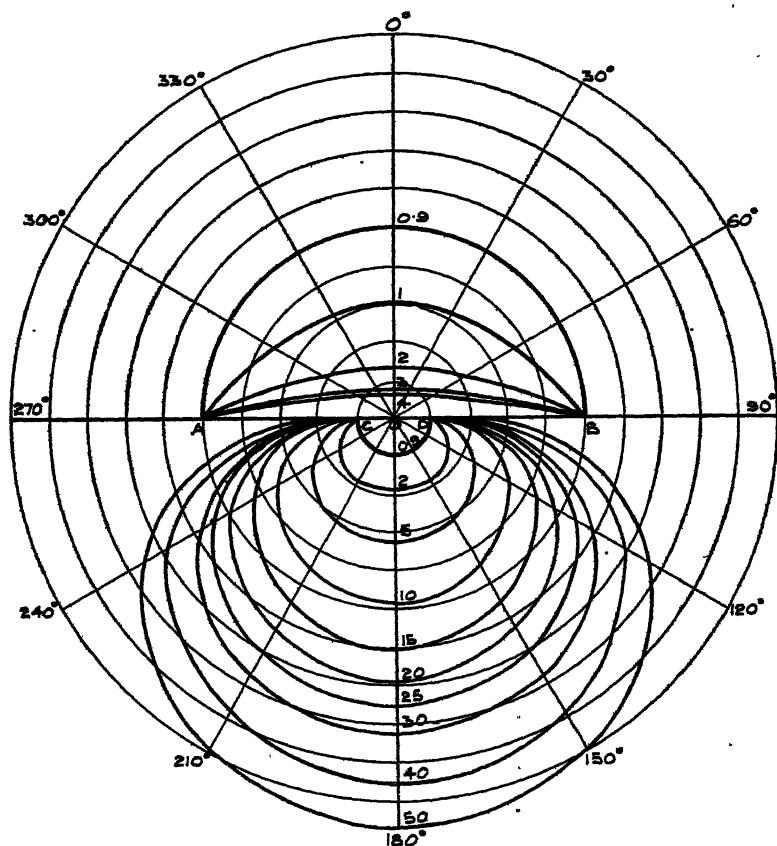
Figure 1 *a*. Curves of equal maximum possible error in determining the position of a transmitter from bearings of maximum error $\pm 0{\cdot}5°$.

The figures on the curves indicate the maximum error in position expressed as a percentage of the length of the base line.

AOB=base line for upper half, i.e. for distances from O less than OA.
COD=base line for lower half, i.e. for distances from O greater than OC.

location will be constant in value. It should be pointed out that these curves become unreliable when the transmitter is in such a position that the bearings from the two direction-finders are nearly equal, i.e., when the transmitter is approximately on the base-line or an extension of it, or when the balloon is distant from the base-line.

In the practical application of the method, observations are made continuously on the signal radiated from the balloon transmitter, and consequently the errors

of individual readings can be smoothed out, to a large extent, by averaging the results over fixed time intervals.   This is an important point to be borne in mind when this method of locating the free balloon is applied to the determination of wind velocity.
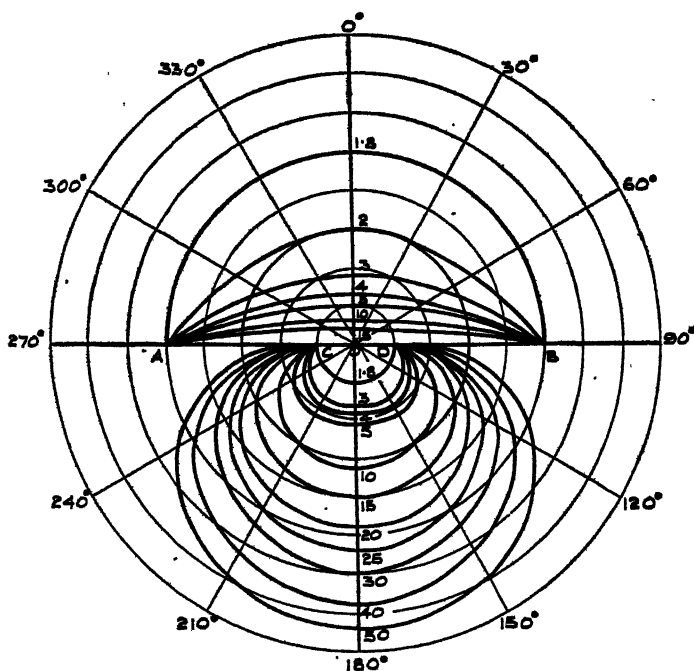
Figure 1 b.   Curves of equal maximum possible error in determining the position of a transmitter from bearings of maximum error ±1°.

The figures on the curves indicate the maximum error in position expressed as a percentage of the length of the base line.

AOB=Base line for upper half, i.e. for distances from O less than OA.
COD=base line for lower half, i.e. for distances from O greater than OC.

§3. PRELIMINARY INVESTIGATION OF DIRECTION FINDING ON BALLOON TRANSMITTERS

As it appeared likely that the two-direction-finder method was the only practicable one available, experimental work was conducted with a view to determining the accuracy attainable with existing ultra-short-wave direction-finders. A previous paper (Smith-Rosé and Hopkins, 1938) has outlined the results obtainable with transportable direction finders on signals emanating from a fixed ground-transmitting station, but it was now desired to ascertain the reliability of the bearings observed on the radiation from a transmitter attached to a free balloon.   Some of the experiments were carried out with the transportable apparatus just referred to, while for later work use was made of fixed-station direction-finding apparatus, the rotating-loop type of which has already been described elsewhere (Hopkins, 1938).

The balloon transmitters used in these investigations were of the type developed at the National Physical Laboratory for meteorological sounding of the

atmosphere, and a description of the transmitter and the experimental work con-
ducted with it has been published by H. A. Thomas (1938). The experiments
were made by releasing a balloon with transmitter attached, and then following the
course of the balloon with a theodolite while bearings were being taken on the
transmitter with the direction-finder. A comparison between the visual and radio
bearings then gave the accuracy of the latter. The radio bearings were taken by
the usual aural-null method, some ten to twenty observations being made per
minute. The mean of such a group of bearings was then compared with the
mean of the theodolite bearings for the same minute, and the summaries given
below refer to these mean observations. The radio observations were sometimes
continued after the balloon had passed out of sight, in order to provide data as to
the range of usefulness of the direction-finder.

(a) *Measurements with field direction-finders.*

The first series of measurements were carried out with the experimental
transportable direction-finders referred to above, and especially developed for
field investigation work. These were of two types : one was fitted with a simple
rotating loop, while the other used a rotating spaced aerial of the Adcock pattern.
The great advantage of using these instruments in the early work was that the
experiments could be conducted at open field sites where previous calibrations
had indicated that errors due to the site itself were likely to be small. As the investi-
gation progressed, it became increasingly evident that the influence of the site, or
the radio interfering effect of any obstacles thereon, is likely to be the limiting
factor in the accuracy of this method of locating elevated radio transmitters
whether attached to balloons or aircraft.

Employing balloon transmitters designed to operate in the wave-length range
8 to 10 metres, the sensitivity of the loop direction-finder was found to be
adequate to enable good bearings to be taken at ranges up to some 40 miles ;
while, with some sacrifice of quality, radio observations have been obtained at
ranges up to about 90 miles. The form of the Adcock direction-finder used in
these early experiments had a maximum range of 20 miles. Later, a more
sensitive Adcock set has been developed and a description of this is in course of
publication. A direct comparison carried out between the loop and the Adcock
direction-finders showed that the means of each group of bearings taken over a
period of one minute were in fairly good agreement, although the oscillation of the
individual bearings due to the swinging of the transmitter aerial was in general
much greater with the loop set than with the Adcock receiver. The results
obtained during balloon ascents in which theodolite observations were taken are
given in table 1. These measurements were made with the direction-finders
placed at several sites ; with the exception of those at which balloons Nos. 4 and
30 were released, these sites were carefully chosen so as to minimize any influence
of the surroundings on the radio bearings.

Consideration of this table shows that in the case of the loop receiver the
extreme error was $7°·5$, while with the Adcock receiver the extreme error was $3°·0$.
It should be noted that the results above cover a range of elevation from $4°$ to $36°$.
The importance of the angle of elevation of the balloon is discussed later, but it
should be pointed out that as the angle of elevation increases, the wanted pick-up,

that is, the e.m.f. induced in the vertical portions of the aerial array, falls off very rapidly. Furthermore, the magnitude of the unwanted pick-up, that is, the e.m.f. induced in the direction-finding system by the swinging of the transmitter aerial from the vertical, increases. Thus at high angles of elevation the accuracy of D.F. observation falls off considerably. The principle of the Adcock direction-finder eliminates, to a large extent, the unwanted pick-up, and this type of direction-finder would therefore be expected to give better results than the loop at high angles of incidence. Unfortunately, the sensitivity of the early experimental Adcock was insufficient to make use of this advantage, and the satisfactory working

Table 1. Results of D.F. tests at field sites with balloons using wave-lengths between 8 and 10 metres

Balloon No.	Duration of visual observations (min.)	Range of elevation during test (deg.)	Estimated distance of balloon at last observation (miles)	Accuracy of D.F. bearings for each minute			
				Extreme error (deg.)	Mean error (deg.)	% within 2°	% within 1°
		Loop results					
4	15	16–21	5	6·0	3·2	50	12
5	7	17–26	3	1·1	0·4	100	86
17	57	12–26	15	2·0	0·9	100	49
18	31	4–12	4	1·7	0·6	100	90
30	12	24–36	5	5·0	1·9	67	42
32	9	28–32	3	7·5	2·9	45	44
		Comprising all the six tests			1·3	88	58
		Adcock results					
17	57	12–26	15	2·0	0·8	100	67
18	31	4–12	4	1·4	0·5	100	96
27	7	21–28	2	1·3	0·7	100	71
28	11	8–16	4	1·0	0·4	100	100
30	12	24–36	5	1·1	0·4	100	92
32	9	28–32	3	3·0	1·2	78	44
38	21	8–17	16	1·6	0·5	100	86
		Comprising all the seven tests			0·6	99	79

range of the instrument even at low angles of elevation was limited to about ten miles. Later work with the more sensitive instruments has shown them to be free from this limitation, and it is not considered that the accuracy of the D.F. bearings will deteriorate appreciably for ranges of the balloon somewhat beyond that at which it was possible to make visual observation during these tests.

The results of this first series of experiments were very encouraging in showing that it was practicable to use an ultra-short-wave direction-finder to determine to a reasonable accuracy the bearing of a free balloon carrying a radio transmitter. The tests showed that with a rotating loop set, 88% of the bearings were correct to within 2°, while with the experimental Adcock sets the accuracy was appreciably higher.

It has naturally been impossible to check the accuracy of the direction-finders when the transmission path has been partly through clouds.* It is, however, reasonable to suppose that clouds will not materially affect these results, except possibly during thundery weather, when ionization in the clouds may be considerable.

It was next decided to ascertain to what extent this accuracy could be improved by using a fixed-station instrument of more rigid and robust design. At the period under consideration, certain technical difficulties were being experienced with the Adcock or spaced-aerial direction-finder, and while these were being investigated, the experimental work with balloon transmitters was continued with the improved form of rotating-loop instrument specially developed for the purpose (Hopkins, 1938).

(b) *Measurements with a direction-finder in a wooden hut*

Preliminary experiments carried out with the improved rotating-loop direction-finder installed in a hut at the Radio Research Station, Slough, indicated that a considerable improvement in instrumental performance was now obtainable. The results of two balloon ascents from Slough in which comparison observations were made with a theodolite are shown in table 2.

Table 2. Results of D.F. tests at Radio Research Station, Slough, with balloons using wave-lengths between 8 and 10 metres

Balloon No.	Duration of visual observations (min.)	Range of elevation during test (deg.)	Estimated distance of balloon at last observation (miles)	Accuracy of D.F. bearings for each minute			
				Extreme error (deg.)	Mean error (deg.)	% within 2°	% within 1°
36	7	20–25	8	3·7	1·1	86	57
37	26	16–26	20	2·9	1·0	89	65
Mean of the two tests					1·0	88	64

These results, which are rather better than those obtained with the field sets, confirmed the view that the effect of the hut on the bearing accuracy would be small. Experiments with a portable transmitter set up close to the hut had already shown that the latter was not effective in producing bearing errors in this wave-band when dry, uniformly wet, or even when one or more sides were dry and the remainder wet.

All these results were sufficiently encouraging to warrant the continuation of the work with loop direction-finders of the fixed-station type on better sites than had so far been available, and the details of this work are described in the following section.

* Modern radar methods are now available to assist in such investigations.

§ 4. DIRECTION FINDING ON BALLOON TRANSMITTERS AT FIXED GROUND STATIONS

(a) *Choice of sites for experimental work*

Through the co-operation of the Meteorological Office, Air Ministry, two direction-finding sites near Larkhill, on Salisbury Plain, were made available for experimental work. Each of these sites was selected by personal inspection and was quite evidently superior to those at present available at the Radio Research Station.

The first site, at Elston Hill, was situated in the centre of a flat expanse of arable land, which was about three-quarters of a mile square, and was reasonably level over the whole area. The hut containing the direction-finder was situated near a footpath at approximately 300 yards from a minor road, and the nearest obstacles of any magnitude were the trees surrounding a farm-house some 1000 yards distant. In the earlier experiments an area of radius 25 yards around the hut was kept free of crops, but this was found to be insufficient, and the radius was later increased.

The other site was at Lavington, about five miles away, and was a large stretch of pasture land reasonably flat, but with a gentle slope of about 1° in one direction and a somewhat steeper slope after a distance of 500 yards in the opposite direction. The nearest obstacles were trees more than 1000 yards away and a steel water tower, also at about 1000 yards.

At each site the direction-finder was housed in a large wooden hut containing no metal-work larger than six inches, and, apart from the instrument, no metal objects formed part of the equipment.

(b) *Calibration of sites*

In order to ascertain the inherent accuracy of each instrument as installed in its hut, each direction-finder was calibrated by observing the bearings on a small portable transmitter set in turn at various positions round the hut at a radius of about twenty yards. In all except the preliminary experiments, care was taken to ensure that the assistant placing the transmitter in position did not produce spurious readings on the direction-finder ; it was found that if he retired to a distance of about ten yards behind the transmitter on the line joining this to the receiver, no error was introduced. To facilitate the repetition of the calibrations at various times, it was found convenient to install a circle of wooden pegs marking the transmitter positions : in the earlier work these were provided at intervals of 30°, but as the work progressed it was found desirable to place additional pegs at 10° intervals. Also in the earlier tests, a landing-type magnetic compass was used to ascertain the true bearing of the various pegs, but various comparisons between this compass and theodolite indicated that the compass readings were limited to an accuracy of about ±0°·3. As the results soon showed that an accuracy of this order was attainable with the direction-finder, the peg positions were finally located to a high accuracy by a theodolite method. The majority of the balloon work at these sites was carried out on a wave-length of 8·5 metres (frequency 34 Mc/s.) arbitrarily chosen, but it was known that ultimately the technique would have to be transferred to the recently allocated international band of 10·7 to 10·9 metres (27·5 to 28·0 Mc/s.). In the calibration work, therefore, the observations

were carried out for a range of frequencies, and it was usually found that the observed bearing was in error and the value of the latter varied with frequency.

Throughout the calibrations, the observed bearings on the local transmitter were sharp, with a swing of about $\pm 0°·2$. As a result, the observational accuracy is considered to be of the order of $0°·1$. In general, the maximum error obtained in the course of the calibrations was about 1°, although in certain circumstances errors of 2° have been noted.

Typical calibration curves obtained for the Elston and Lavington sites in September and December 1938 on a frequency of 34 Mc/s. are shown in figures 2 *a* and 2 *b*. It is seen that there is an appreciable change in the calibration with time, the
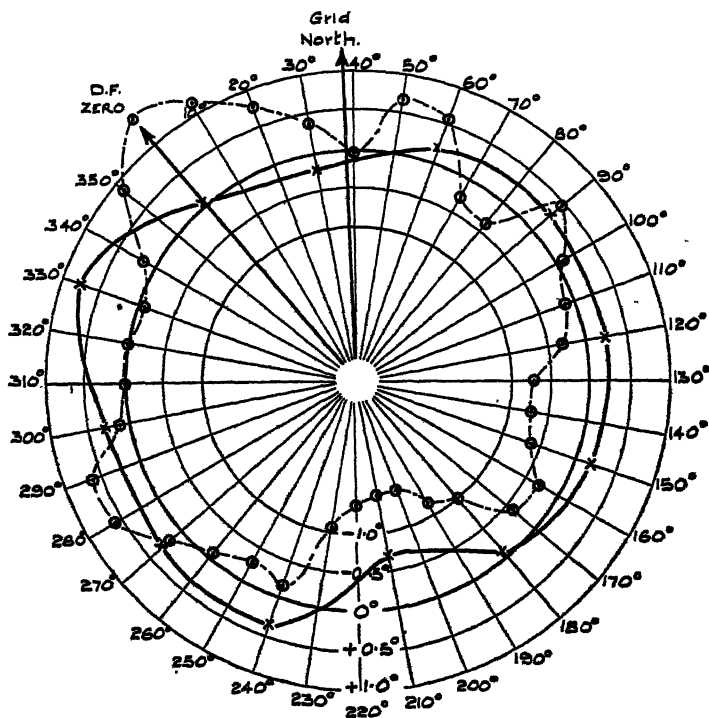
Figure 2 *a*. Site calibrations with local transmitter on 34 Mc/s.
Elston Hill site.

Error $\begin{cases} \times & \text{September 1938.} \\ \odot & \text{December 1938.} \end{cases}$

change being more marked at Elston Hill (figure 2 *a*) than at Lavington (figure 2 *b*). A maximum of about $0°·2$ of this change was subsequently traced to a possible change in the position of the flexible transmission line leading from the turn-table down to the receiver on the floor. This fault in the instrument has now been rectified.

On a few occasions, calibrations of a similar nature to the above have been conducted with the portable transmitter removed to distances up to 150 yards. It was found that the errors in the observed bearings varied within the range $\pm ·0°3$, depending upon both the actual azimuth and the working frequency. It has not so far been practicable to extend this type of measurement to greater distances with a view to ascertaining the degree to which the short-radius calibration curves

may be applied as corrections to bearings observed on distant transmitters. In view of this practical difficulty, the applicability of such ground calibrations to results obtained on balloon transmitters will depend upon the relative importance of errors having their cause only a few yards from the direction finder and those
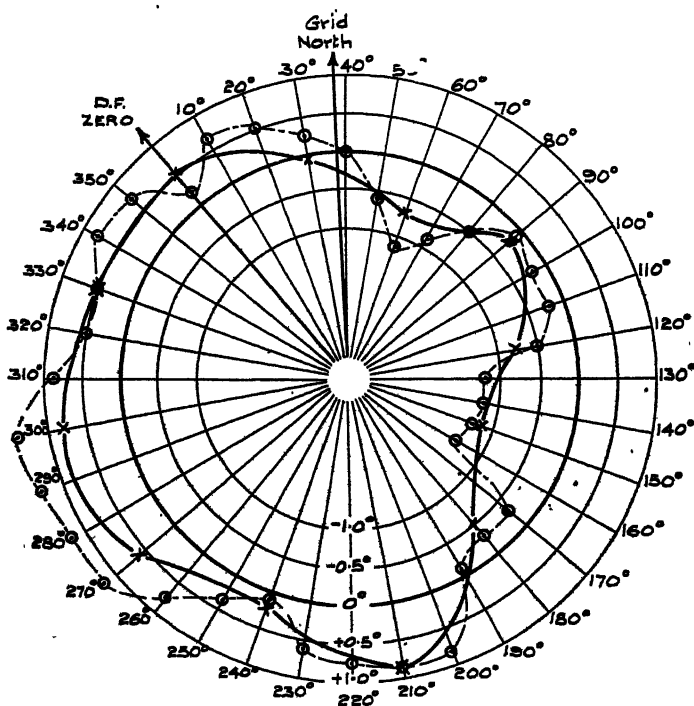
Figure 2 *b*. Site calibrations with local transmitter on 34 Mc/s. Lavington site.

Error $\begin{cases} \times & \text{September 1938.} \\ \odot & \text{December 1938.} \end{cases}$

caused by obstacles farther away. It was with the object of throwing more light on this aspect of the problem that further experiments were made at the Elston and Lavington sites, in which the radio bearings were directly compared with visual bearings obtained with theodolites.

(c) *Further results on balloon transmitters*

In these experiments a number of balloons equipped with transmitters were released from each of the two sites, and the course of each of these was followed visually with a theodolite set up near the direction-finder. A second theodolite was used to determine the position of the balloon during the early part of each ascent, so that the observations on the first theodolite could be corrected for the small error introduced by the distance which separated the direction-finder from the theodolite. During the ascent of a balloon, radio-bearing observations (on a wave-length of 8·5 metres) were taken continuously at the rate of some 10 to 20 per minute, and theodolite observations were taken at half-minute intervals. In this way it was possible to compare directly the mean bearing of the

balloon transmitter during each minute as given by the direction-finder with the mean true bearing as given by the theodolite results. A summary of the results obtained during twelve ascents at Elston and nine at Lavington is given in table 3. These results refer to the actual readings on the instruments, before introducing any corrections derived from the ground calibrations, and they may therefore be compared directly with the earlier results summarized in table 1.

Table 3. Results of tests with radio-equipped balloons using direction-finders installed in wooden huts ($\lambda = 8\cdot 5$m.)

Balloon No.	Duration of visual observa-tions (min.)	Range of elevation during test (deg.)	Estimated distance of balloon at last observation (miles)	Accuracy of D.F. bearings for each minute				
				Extreme error (deg.)	Mean error (deg.)	% within 2°	% within 1°	% within 0°·5
(a) Direction-finder at Elston Hill								
1	6	41–45	1	1·6	0·8	100	60	60
2	43	23–60	11	5·0	1·5	83	31	21
4	5	9–61	1	3·5	1·6	80	40	0
6	54	11–42	18	5·7	0·7	96	78	63
8	73	10–24	35	2·0	0·6	99	78	49
10	73	13–25	19	4·3	0·8	99	71	18
11	60	12–21	22	4·7	0·7	98	93	26
12	7	—	1	9·3	2·3	57	57	29
14	7	11–22	2	0·6	0·4	100	100	57
16	29	7–12	15	2·5	0·4	95	95	89
19	20	19–53	4	2·0	1·1	100	47	0
22	28	20–34	9	4·2	1·0	93	43	29
Comprising all the twelve tests				0·8		96	72	37
(b) Direction-finder at Lavington								
3	28	7–18	6	5·0	0·9	89	68	64
5	35	8–24	11	1·1	1·7	100	86	23
7	34	17–27	8	5·2	0·5	94	94	79
9	8	28–45	2	2·1	1·0	86	57	14
13	8	6–23	1	9·6	2·1	62	62	50
15	17	7–13	8	1·7	0·6	100	88	47
17	16	8–23	3	1·1	0·4	100	94	56
20	35	9–21	10	4·4	0·5	94	94	77
59	68	6–29	9	4·9	0·6	97	82	50
Comprising all the nine tests				0·7		95	84	55

It will be seen that this improved form of loop direction-finder erected on a superior site gives a considerable increase in overall accuracy in this work. The maximum error in the radio bearings was about 5°, while about 95% are within 2° and nearly 80% within 1°. It was found that the largest errors in bearing generally occur in the first three minutes of flight. Factors contributing to this effect are the violent aerial swings in this period, the observer's initial "mental inertia" in following the consequent variations in bearing, his preoccupation with tuning adjustments, and the high rate of change of mean bearing in the first minute or so

of flight.   These large initial errors are, therefore, not truly representative of the capabilities of the direction-finders observing on a balloon transmitter during the major portion of a flight.

In order to ascertain to what extent the ground calibration may be usefully applied to these results, a direct comparison has been made of the errors in bearings before and after the appropriate correction derived from the site calibrations at twenty yards radius was applied to each D.F. observation (minute means).   This comparison is shown in the following table, which refers to the whole of the results given in table 3, omitting the first three minutes of each ascent, and for both the Elston and Lavington sites.

This table clearly demonstrates that the ground calibration made at a short distance around each direction-finder is quite useful in improving the overall accuracy of the observed radio bearings on balloon transmitters.   It is evident, therefore, that at least part of the error in bearings is associated with the immediate vicinity of the direction-finding installation: the instrument itself and the hut would appear to be exonerated by the irregular shape of the calibration curves on

Table 4.   Effect of correction for ground calibration on bearings taken on radio balloon transmitters

Site	Correction for ground calibration applied	Percentage of minute means of radio bearings differing from visual bearings by not more than			
		2°	1°	0°·6	0°·2
Elston	· No	99	76	49	14
,,	Yes	99	91	73	33
Lavington	No	99	90	68	23
,,	Yes	100	95	82	44

any one frequency.   Reference may be made at this stage to the high overall accuracy which is now associated with these direction-finding arrangements. Bearing in mind that the results refer particularly to the mean observed radio bearings over one-minute intervals during the whole flight of a number of free balloons, it is to be noted that over 90% of these mean bearings are correct to within 1°, and that the probable error calculated from all the readings, after correction for ground calibration, is 0°·35 for Elston and 0°·25 for Lavington.

One further aspect of the results is illustrated by table 5, in which the errors of the radio bearings are separated for various arcs of elevation covered by the balloons in flight.   The first three minutes of each ascent have been neglected, for reasons explained above.

From this table it is seen that there is a definite tendency for the errors to increase with the angle of elevation.   This is to be expected from the fact that with a swaying aerial the effect of the horizontally polarized component in the emitted radiation will be to produce errors on the loop direction-finder which increase with the angle of elevation.   So long as this work is restricted to loop sets it will obviously be advantageous to arrange the balloon ascents to give as small angles of elevation as possible, but this is not always practicable when a rapid

ascent to high altitudes is required. The conversion of the direction finders to the spaced-aerial or Adcock system should, therefore, give a considerable improvement in the radio-bearing accuracy from this point of view.* In addition, the use of the Adcock system brings about a greater steadiness of the individual observations, the variations of which are due to the swaying of the balloon transmitter aerial.

§ 5. DISCUSSION OF RESULTS AND CONCLUSIONS

Before summarizing the conclusions which may be drawn from the results of the experiments described above, it will be useful to review the various types of error to which the bearings observed on a radio direction finder are subject. It is essential to understand clearly these various errors and their relative importance and to investigate their causes before further progress can be made

Table 5. Variation of mean error of radio bearings with elevation of transmitter, after correcting for local calibration

Site	Arc of elevation (degrees)	Mean arithmetic error (degrees)
Elston	5–10	0·2
	10–15	0·3
	15–20	0·4
	20–25	0·5
	25–30	0·7
	30–35	0·7
	35–40	0·9
	40–45	1·3
Lavington	5–10	0·3
	10–15	0·3
	15–20	0·4
	20–25	0·4
	25–30	0·6
	30–35	0·6

in the particular application of direction finding discussed in this paper, viz., the location of a radio transmitter attached to a free balloon. It will be evident that the conclusions may equally apply to the case of the location of aircraft transmitters within the limitations of distance, type of aerial, etc., imposed by the former application.

(a) *General discussion of errors in direction finding*

In general direction-finding work it is convenient to classify the various effects observed under the headings of *Instrumental* (including reciprocal effects and polarization errors), *Site, Ground path, Atmospheric path* and *Source* errors.

* With spaced-aerial systems the amplitude of the wanted voltage due to the vertically polarized component from the balloon is likely to fall off more rapidly with increase of elevation angle than that due to a similarly polarized reflection from an obstacle near the site. Even with an ideal Adcock system, therefore, we should expect site errors to be more troublesome at high angles of elevation.

Instrumental errors are inherently due to defects in the design or construction of the instrument, which introduce an error into the bearings indicated or observed.   The reciprocal-error effect is made evident by the fact that the observed bearings in the two possible positions of the rotating system are not exactly 180° apart.   Although the resulting error in bearing can be reduced by taking the mean of such readings, with the appropriate 180° correction, it is desirable that the reciprocal error should be reduced as much as possible.   The effect arises from lack of symmetry of the receiving aerial circuit, and can be reduced or avoided by careful design and construction supplemented, if necessary, by a compensating device.   It is not usually difficult to reduce such an effect to less than 0°·5, and in the instruments used in the later portions of this investigation the error was reduced to less than 0°·1.

The polarization errors are due to the reception of horizontally polarized electric waves on the aerial system, the magnitude of the effect decreasing as the angle of incidence of the arriving waves is increased, to a minimum for horizontally propagated waves.   The loop direction-finder is very susceptible to such effects, but the magnitude of the error depends upon the relative intensities of the vertical and horizontal fields in the incoming waves.   Using a balloon transmitter with a steady vertical aerial, no such polarization errors would be experienced on the loop direction-finder.   In practice, however, the transmitting aerial performs a conical oscillation about an axis which, due to the horizontal motion of the balloon, may not be vertical; and it is therefore to be expected that variable polarization errors would be experienced with such a direction finder.   The results already obtained and discussed in the previous sections show that by taking the mean of several bearings observed over one-minute periods the effect of such oscillation of the bearing can be kept reasonably small, providing the angle of elevation of the balloon itself is not excessive.

Site errors arise from the reception of re-radiated fields from currents induced by the incoming waves in obstacles such as trees, buildings and metal conductors on the site of the direction finder which may arbitrarily be taken as the area of terrain within about a quarter of a mile from the direction finder.   From earlier general experience in direction finding on other wave-lengths, it is known how to avoid the larger errors due to bad site conditions, but with the general improvement in the accuracy of direction-finding technique, the matter has called for further investigation on a wide scale.   The subject is found to be of a somewhat complex nature, and detailed discussion of the general problem is beyond the scope of the present paper.   The results of experiments described above indicate the order of the effects observed in relation to the particular application dealt with in this investigation.

Ground and *atmospheric path* errors are due to any deviation or reflection effects introduced in the propagation of the waves and caused either by topographical features, such as hills, masses of trees or the ground, or by atmospheric conditions such as clouds or temperature gradients in the path between transmitter and receiver.   The experiments described in this paper tend to confirm the view that the error due to these effects of topography—although small—are likely to be a factor limiting the accuracy of bearings taken on signals from balloon-borne transmitters.   On the wave-lengths and at the distances used for

this work it is considered that the field at the direction finder is free from radiation returned from the ionosphere.

Source error arises from any peculiarities in the distribution of radiation from the immediate vicinity of the transmitter, due either to the aerial system employed or to re-radiation effects from obstacles close to the transmitter. In the case of free balloons of the type used in this investigation, such source effects are quite negligible.

(b) *Conclusions*

The above general discussion makes it clear that in order to obtain a high degree of accuracy in locating free balloons by radio direction-finding methods, it is necessary to pay special attention to the instrument itself and the site on which it is placed.

The greater part of the investigation described in this paper has been carried out with direction-finders which have been specially developed to have a high instrumental accuracy of bearing indication, which is of the order of $0°\cdot1$ under the best conditions, using a working wave-length of about 10 metres. Some difficulty has been experienced in utilizing this accuracy, however, due first to effects introduced by the site on which the direction finder is placed, and secondly, by the departure during flight of the balloon aerial from a vertical position. By a careful choice of the site on which the direction-finder is placed, the former error has been restricted to within a range of about $\pm 1°$, but variations of about half this magnitude have been experienced in the course of successive calibrations with a three-monthly interval. The cause of this error and of its variation have not so far been ascertained, but the results of various experiments suggest that the bearings observed on a balloon transmitter can be improved in accuracy by applying a correction for this local site error. The effects of the swaying of the transmitting aerial while a balloon is in flight can be largely compensated for by taking the mean of a succession of radio-bearing observations over a period of a minute, and the results of tests carried out on some twenty balloon flights under these conditions show that 95% of the observed bearings are correct to within 2°, and nearly 80% correct to within 1°. These results can be still further improved with the use of a more fully developed Adcock or spaced-aerial direction-finder than was available for the experiments described in this paper.

It must be pointed out that since the assessment of the direction-finding accuracy is carried out by comparing the radio observations with visual bearings obtained by means of theodolites, all the tests described above were necessarily conducted under fair-weather conditions when visibility was reasonably good. In many cases, however, the radio observations were continued long after the balloon had passed out of sight above the clouds, and the general experience obtained under such conditions indicates that the above accuracies are likely to be maintained for the whole period during which the signals from the transmitter are of adequate strength at the direction-finding receiver. It thus appears that within the range of such balloon location work, which may cover distances up to the order of 50 miles and altitudes up to some 40,000 feet, no serious errors due to the propagation of waves through the atmosphere are likely to be encountered.

The work described above is limited to a study of the accuracies of individual radio direction-finders. The resulting accuracies in the location of a free balloon

by means of two direction-finders operating simultaneously can be deduced from theoretical curves given in the paper. The extension of the experimental investigation to such simultaneous operation and the resulting application for meteorological purposes must remain outside the scope of the present paper.

ACKNOWLEDGMENTS

The investigations described above were carried out for the Director of the Meteorological Office, by whose permission the paper is now published. The authors are particularly indebted to Mr. L. G. Hemens, B.Sc., and the staff of the Meteorological Office at Larkhill, with whose co-operation the work at Larkhill was conducted. They also have to acknowledge the assistance rendered by Mr. W. Dougharty, B.Sc., throughout the experiments.

REFERENCES

CORRIEZ and PERLOT, 1935. "The radiogoniometric method of the National Bureau of Meteorology. for measuring the direction and velocity of winds during overcast weather." *La Météorologie*, August, 368.

DIAMOND, H., HINMAN, W. S. and DUNMORE, F. W., 1937. "The development of a radio meteorograph system for the Navy Department." *Bull. Amer. Meteorological Soc.* **18**, 95.

HOPKINS, H. G., 1938. "A loop direction finder for ultra-short waves—wave-range 6 to 11 metres." *Wireless Engr.* **15**, 651.

SMITH-ROSE, R. L. and HOPKINS, H. G., 1938. "Radio direction finding on wavelengths between 6 and 10 metres (frequencies 50 to 30 Mc/s.)" *J. Inst. Elect. Engrs.* **83**, 87.

THOMAS, H. A., 1938. "The determination of the meteorological conditions of the atmosphere by the use of radio-sounding balloons." *Proc. Roy. Soc.* A, **167**, 227.

THOMAS, H. A. "A method of determining the position of an elevated radio transmitter." British Patent 28,921/37.

THE ACCURATE DETERMINATION OF CELL DIMENSIONS FROM SINGLE-CRYSTAL X-RAY PHOTOGRAPHS

BY MARGARET C. M. FARQUHAR AND H. LIPSON,

Cavendish Laboratory, Cambridge

ABSTRACT. The importance of improving the accuracy of the determination of cell dimensions determined from single-crystal photographs is pointed out and the general principles upon which such improvements can be obtained are outlined. These principles, which are based on those used for powder photographs, are illustrated by application to an orthorhombic crystal, thallium hydrogen tartrate, the cell dimensions being determined to an accuracy of the order of 0·005 %. Extensions of the method to monoclinic and triclinic crystals are also discussed.

§ 1. INTRODUCTION

ROTATION and oscillation photographs provide a ready means for finding the unit-cell dimensions of a crystal, since, if the crystal is rotating about a crystallographic axis, the distances apart of the layer lines provide a measure of one of the lattice spacings. The accuracy of this method is, however, not high;

it is great enough for unequivocal indexing of the spots on oscillation photographs, but few workers would claim an accuracy of more than about 1% for cell dimensions deduced in this way. This accuracy is probably sufficient for most crystal-structure work, since interatomic distances are not usually determined to this accuracy. But recent work with three-dimensional Fourier methods suggests that an accuracy of 1% in interatomic distances has been approached (Jeffrey, 1945) and even surpassed (Booth, 1945), and therefore it appears that for crystals studied by these methods a greater accuracy in the determination of cell dimensions would be desirable.

In certain other applications of x-ray diffraction an increased accuracy would also be desirable. For example, in the determination of molecular weights an accuracy of 0·03% is needed to match the 0·1% that can sometimes be obtained in the determination of densities. Buerger (1937) has described a special Weissenberg instrument by which this accuracy can be obtained. It is, however, unnecessary to have a special camera for the purpose; an accuracy of 0·01% can easily be obtained with an ordinary single-crystal goniometer, only small changes in design being necessary. The only qualification is that the crystal should be perfect enough to give sharp spots with Bragg angles θ near 90°.

§2. THEORY OF THE METHOD

The method to be described is based essentially on that already extensively used for powder photographs (Lipson and Wilson, 1941). The positions of re-solved α doublets with θ near 90° are measured and the value of one of the cell dimensions is deduced from each; the values are plotted against $\sin^2\theta$ and the value at $\sin^2\theta = 1$ is taken as correct.
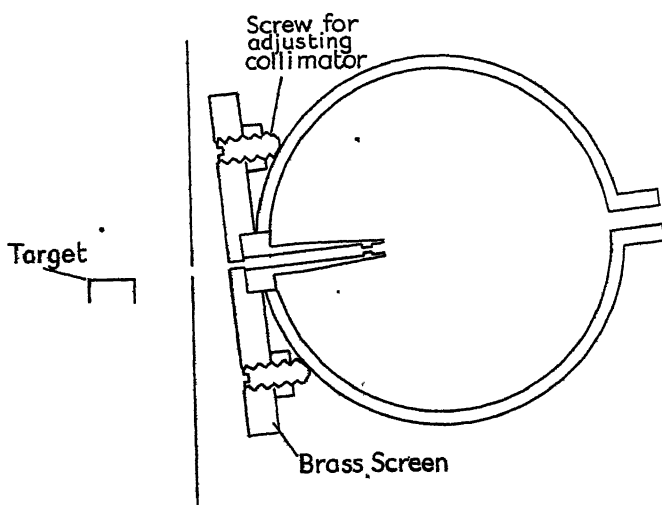
Figure 1 *a*. Section of camera with collimator attached.

A difficulty arises in the determination of the Bragg angles of the spots. Single-crystal cameras are not usually fitted with calibrated knife-edges, and to introduce these would require considerable redesign. This difficulty can be avoided by using the van Arkel method (Lipson and Wilson, 1941) of mounting the film; the

x rays enter through a hole in the film and the reflections with θ near $90°$ are then recorded near the middle of the film, not near the edges. (This film mounting is not recommended for ordinary indexing, where it is important to know the position corresponding to $\theta = 0°$). The angles can then be evaluated with a knowledge of the camera radius, and even though this may not be the effective radius of the film, because of film shrinkage, the error is eliminated by the method of extrapolation.
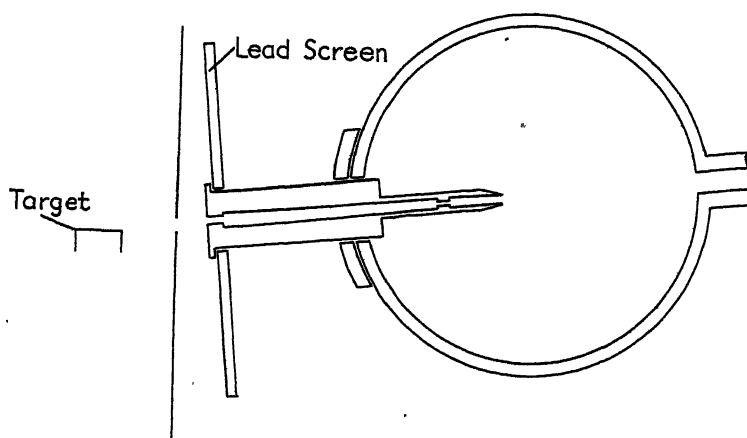
Figure 1 *b*.   Section of 3-cm. radius Unicam camera.

The film can be most easily mounted in the camera if the collimator is rigidly attached to the camera, since then only a small hole is needed in the film, as shown in figure 1 *a*. This has an additional advantage over the more usual arrangement, shown in figure 1 *b*, in that the camera can be put closer to the x-ray tube. With a camera of 3-cm. radius it was found that this modification resulted in about a three-fold increase in intensity of the spots in the film. The presence of the collimator in the camera also facilitates the loading of the film, and renders unnecessary the movement of the goniometer between exposures.

§3. USE OF THE METHOD FOR AN ORTHORHOMBIC CRYSTAL

The spots which are used are those on the zero layer line of an oscillation photograph. If an axis of the crystal is oscillating symmetrically with respect to the x-ray beam a symmetrical photograph will be obtained and values of θ can be derived from the distances S between corresponding spots on the two sides of the film. If S is the distance between two such spots,

$$90° - \theta = \frac{S}{4R} \cdot \frac{180}{\pi}. \qquad \ldots\ldots(1)$$

For a camera of 3-cm. radius this reduces to

$$90° - \theta = 4 \cdot 77S. \qquad \ldots\ldots(2)$$

The symmetrical oscillation photograph provides another advantage. In figure 2 is shown the $a^*\,b^*$ section of an orthorhombic reciprocal lattice with the reflecting circles (Bernal, 1926) symmetrically disposed with respect to the b axis. It will be seen that the spots with θ near $90°$ have a large k and small h index. These spots will be particularly well suited to the determination of the value of b,

since a small error in *a* will not greatly affect the results. In order to find *a*, another photograph can be taken with the *a* axis oscillating symmetrically with
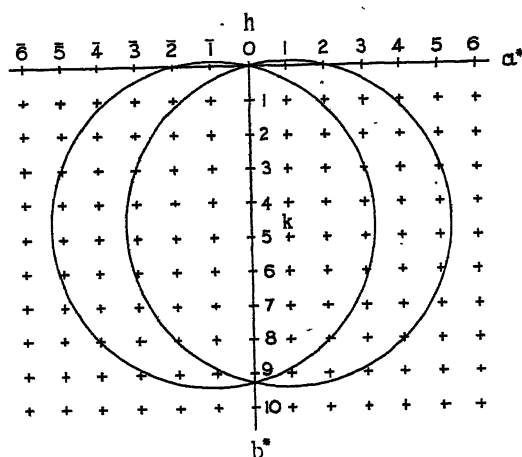
Figure 2. Section of orthorhombic reciprocal lattice and extreme limits of circles of reflection for a symmetrical oscillation photograph.

respect to the x-ray beam, the axis of rotation being the same as for the former photograph. To determine the value of *c*, however, either the *a* or *b* axis must be set as axis of rotation.

§4. A PRACTICAL EXAMPLE

The method has been used to determine the cell dimensions of thallium hydrogen tartrate. Groth (1910) gives the axial ratios as

$$a:b:c = 0 \cdot 6976 : 1 : 0 \cdot 7275,$$

and these values are supported by the measurements of Porter (1928), who gives

$$a:b:c = 0 \cdot 6950 : 1 : 0 \cdot 7234.$$

A crystal was mounted with its *a* axis as axis of rotation; this axis was parallel to the longest dimension of the crystal and, on the oscillation photographs taken with Cu Kα radiation, the x-ray reflections were sharp and accurately measurable. Measurements of the layer lines gave $a = 7 \cdot 63$ kx., from which, using the axial ratios given by Groth, we obtain $b = 10 \cdot 94$ kx. and $c = 7 \cdot 96$ kx. These values were used for the analysis of the photographs.

Two symmetrical 15° oscillation photographs were taken, as explained in §3. The positions of all the resolved α doublets on the zero layer line of each photograph were measured; the results for the determination of *c* are given in table 1. The points as plotted (figure 3 *a*) for extrapolation show a great deal of scatter, and it is impossible to deduce more than that the value of *c* is in the neighbourhood of 7·91 kx. The scatter is not, however, a random one; the points for reflections with larger *k* tend to lie above the others. This is an indication that the axial ratio is n error. This was confirmed by the measurement of *b*, which gave $b = 10 \cdot 99$ kx.; the axial ratio is thus about $0 \cdot 7197_5$, not $0 \cdot 7275$. A recalculation of *c* with this axial ratio (shown in the lower part of table) gave the points shown in figure 3 (*b*);

Table 1

The derivation of c from an oscillation photograph of a single crystal of thallium hydrogen tartrate taken with Cu Kα radiation in a camera of radius 3 cm.

$h\ k\ l$	Radn.	θ	cosec θ	$\frac{1}{2}\lambda\sqrt{\frac{c^2}{b^2}k^2+l^2}$	c	$\sin^2\theta$
0 6 8	α_1	62·267	1·129783	7·00548	7·9147	0·784
	α_2	62·496	1·127423	7·02296	7·9278	0·787
0 4 9	α_1	66·832	1·087720	7·27095	7·9088	0·845
	α_2	67·171	1·084991	7·28909	7·9086	0·850
0 5 9	α_1	70·529	1·060660	7·46201	7·9147	0·889
	α_2	70·930	1·058067	7·48063	7·9150	0·893
0 2 10	α_1	79·153	1·018192	7·76792	7·9092	0·965
	α_2	79·892	1·015767	7·78731	7·9101	0·969
0 3 10	α_1	83·894	1·005705	7·86793	7·9128	0·989
0 6 8	α_1			6·98834	7·8953	0·784
	α_2			7·00579	7·8985	0·787
0 4 9	α_1			7·26362	7·9008	0·845
	α_2			7·28175	7·9006	0·850
0 5 9	α_1			7·45085	7·9028	0·889
	α_2			7·46945	7·9032	0·893
0 2 10	α_1			7·76621	7·9075	0·965
	α_2			7·78560	7·9084	0·969
0 3 10	α_1			7·86413	7·9090	0·989
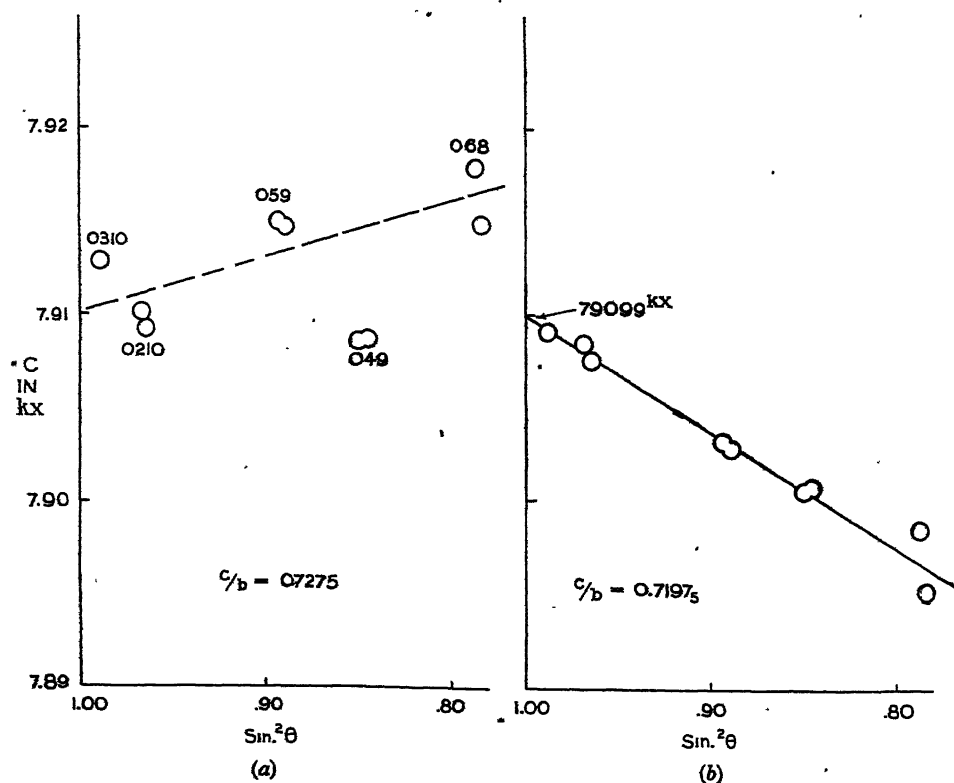
Figure 3. Extrapolation curves for the determination of c: (a) with an inaccurate value of c/b; (b) with an accurate value of c/b.

these lie satisfactorily in a straight line and give the value of c as $7 \cdot 9099 \pm 0 \cdot 0003$ kx. A similar recalculation of b gave $10 \cdot 9945 \pm 0 \cdot 0005$ kx.

In order to find a, a photograph was taken with the crystal set with the b axis as axis of rotation. With $15°$ oscillation and Cu Kα radiation, however, no spots with θ greater than $70°$ appeared on the symmetrical photograph. The range of oscillation was therefore increased to $30°$ by taking two successive exposures on the same film, and two reflections with higher angles were recorded. The results were not so good as with the other photographs, in that the extrapolation curve was not straight, but nevertheless the value of a was found to be $7 \cdot 6395 \pm 0 \cdot 0005$ kx. As a check, the value of c was also deduced from a photograph taken with the same axis of rotation; the value, $7 \cdot 9101 \pm \cdot 0004$ kx., agreed with the previous value well within the limits of experimental error.

To summarize, the results for the crystal are

$$a = 7 \cdot 6395 \text{ kx.,} \qquad b = 10 \cdot 9945 \text{ kx.,} \qquad c = 7 \cdot 9100 \text{ kx.,}$$
$$a : b : c = 0 \cdot 69485 : 1 : 0 \cdot 71945.$$

The lattice parameters are probably correct to $0 \cdot 0005$ kx. and the axial ratios to $0 \cdot 0001$. (It is interesting to note that the value of a deduced from layer-line measurements was correct to $0 \cdot 1 \%$, but this agreement is probably fortuitous.) The density ρ, calculated by means of the formula

$$\rho = 1 \cdot 65026 \, \Sigma A / abc, \qquad \qquad \ldots \ldots (3)$$

where ΣA is the sum of the atomic weights of the atoms in the unit cell (Stockdale, 1940; Foote and Jette, 1940), is $3 \cdot 511_5 \pm 0 \cdot 001$ gm/cc., which is about $0 \cdot 5 \%$ higher than the experimentally determined values of $3 \cdot 491$ and $3 \cdot 496$ gm./cc. (Porter, 1928).

§ 5. DEPENDENCE OF THE RESULTS UPON THE VALUE ASSUMED FOR THE RADIUS OF THE CAMERA

It might be thought that the assumption of a value of 3 cm. for the radius of the camera might lead to appreciable errors since film-shrinkage can introduce errors of the order of $0 \cdot 3 \%$ (Lipson, 1942). The calculation of c, as shown in the lower part of table 1, was therefore repeated with constant $4 \cdot 77$ in equation (2) changed to $4 \cdot 80$; this value was chosen because it gave a much smaller drift in the results. The value of c so deduced was $7 \cdot 9097$ kx., which agrees with the previous value well within the limits of error. This source of error, although it is small, would be eliminated if a camera with calibrated knife-edges were used, but it will be seen that the accuracy obtainable even without such knife-edges is quite high.

§ 6. DEPENDENCE OF THE RESULTS ON THE METHOD OF MEASUREMENT

The films were measured in an instrument of the type described by Sears and Turner (1941), and the positions of strong spots could be read to an accuracy of about $0 \cdot 03$ mm. Since such an instrument may not be available in some laboratories, it was thought desirable to find out how much accuracy would be lost if much simpler methods of measurement were used. Accordingly the spots on the photograph that forms the subject of table 1 were measured by visual

comparison with a steel scale graduated in $\frac{1}{2}$ mm. The result was $7 \cdot 909 \pm 0 \cdot 002$ kx. Thus even with this simple method of measurement an accuracy of about $0 \cdot 02\%$ can be obtained.

§7. EXTENSION TO THE MONOCLINIC AND TRICLINIC SYSTEMS

Similar methods can be used for the determination of $a \sin \beta$, b and $c \sin \beta$ for a monoclinic crystal, for both the a^* b^* and the b^* c^* sections of the reciprocal lattice are orthogonal. The determination of β, however, involves some difficulties, for it has to be obtained from the $h0l$ spectra, the relation being

$$\sin^2 \theta = \tfrac{1}{4}\lambda^2 (h^2 a^{*2} - 2hl a^* c^* \cos \beta - l^2 c^{*2}). \qquad \ldots\ldots(4)$$

If a^* and c^* are known, then β can be derived from spectra for which neither h nor l is zero; for greatest accuracy the spots used to determine β should be those with h and l nearly equal. It is not possible, however, to arrange for a symmetrical photograph except by complete rotation, and it may not be possible to index unambiguously all the spots on a rotation photograph.

Two possibilities of overcoming these difficulties suggest themselves; first, a composite photograph may be taken with a series of small oscillations so that chosen spots appear on both sides of the film; secondly, the crystal may be coated with a fine powder of a crystalline material such as quartz so that the Bragg angles of the spots can be deduced by interpolation from the Bragg angles of the powder reflections. This latter method is less troublesome, although it would mean that the crystal could not be used for any other purpose.

For the triclinic system it is impossible to take symmetrical oscillation photographs, and, therefore, methods such as those described in the preceding paragraph would have to be used for all the lattice parameters.

REFERENCES

BERNAL, J. D., 1926. *Proc. Roy. Soc.* A, **113**, 117.
BOOTH, A. D., 1945. *Nature, Lond.*, **156**, 51.
BUERGER, M. J., 1937. *Z. Kristallogr.* **97**, 433.
FOOTE, F. and JETTE, E. R., 1940. *Phys. Rev.* **58**, 81.
GROTH, P., 1910. *Chemische Krystallogr.* **3**, 322.
JEFFREY, G. A., 1945. *Nature, Lond.*, **156**, 82.
LIPSON, H., 1942. *J. Sci. Instrum.* **19**, 63.
LIPSON, H. and WILSON, A. J. C., 1941. *J. Sci. Instrum.* **18**, 144.
PORTER, M. W., 1928. *Z. Kristallogr.* **68**, 531.
SEARS, J. E. and TURNER, A., 1941. *J. Sci. Instrum.* **18**, 17.
STOCKDALE, D., 1940. *J. Inst. Metals*, **66**, 287.

THE RATIO OF THE NEW INTERNATIONAL LUMEN TO THE LIGHTWATT

By E. F. CALDIN,

Queen's College, Oxford; now at Leeds University

ABSTRACT. The ratio of the new international lumen to the lightwatt is expressed as a function of the radiation constants alone, to a close approximation. Its value is calculated for various recommended values of these constants. The corresponding values of the ratio of the present, or "old", international lumen to the lightwatt are also given.

§ 1. INTRODUCTION

THE ratio of the practical unit of luminous flux (the lumen) to the lightwatt is often needed in calculations on the emission of light from total radiators and other sources. (It is numerically equal to the value, in lumens per watt, of the luminous efficiency of radiation of the wave-length for which the relative luminosity function has its maximum value unity.) It has long been a source of uncertainty, for three reasons. In the first place, the present, or "old", international candle, from which the lumen is derived, is not physically defined. This difficulty will be removed when the new international candle comes into use as the standard of intensity, for this unit is defined in terms of the brightness of a total radiator at the temperature of freezing platinum, and the new lumen will be derived from it. Secondly, the ratio of the new lumen to the lightwatt depends on the form of the relative luminosity function. This is settled if we accept the internationally-agreed table of values which has been in use since 1924. Thirdly, the ratio depends on the values of the radiation constants C_1 and C_2; these have undergone considerable changes in the last twenty-five years and are still somewhat a matter of choice and liable to revision. A direct computation of the ratio for each pair of values of C_1 and C_2 would be very laborious, and a rapid method of computation of sufficient accuracy is clearly desirable. In this paper a close approximation is developed which expresses the ratio of the new international lumen to the lightwatt as an explicit function of C_1 and C_2 alone; the value of the ratio corresponding to any pair of values of the radiation constants can thence be readily calculated. The limits of error are determined mainly by the experimental error in the determination of the melting-point of platinum.

A relation between the temperature and brightness of a total radiator has been given in a previous paper (Caldin, 1945). This relation assumes the C.I.E. values for the relative luminosity function (C.I.E., *Compte Rendu*, 1924; Judd, 1931), and it can be expressed as an explicit function of C_1 and C_2. For a perfect diffuser such as a total radiator, the total luminous flux per unit area of surface is equal to π times the brightness, and so a relation between temperature (°K.) and total luminous flux per unit area (lightwatts cm.$^{-2}$) is easily derived.

Using this relation, it is possible to relate the lightwatt directly to the new inter-national lumen, since the brightness of a total radiator at the freezing point of platinum is by definition equal to 60 new candles per cm² of projected area (*Inter-national Lighting Vocabulary*, C.I.E., 1938), and so the total luminous flux is 60π new lumens per cm² of the area of the source. The value on the international scale of the temperature of freezing platinum is itself dependent upon the value of C_2, since it is measured by finding the ratio of the energy flux at a fixed wave-length due to a total radiator at the temperature of freezing platinum to that at the temperature of freezing gold (Roeser, Caldwell and Wensel, 1931). Thus the total luminous flux per unit area in lightwatts cm.⁻² of a total radiator at the temperature of freezing platinum can be expressed (assuming the C.I.E. relative luminosity function) as an explicit function of C_1 and C_2 alone. For any pair of values of C_1 and C_2, the ratio of 60π to the value obtained gives the ratio of the new lumen to the lightwatt. The ratio of the present, or "old", international candle to the new candle has been experimentally determined (for references see Wensel, 1939, p. 393), and it appears that the brightness of a total radiator at the temperature of freezing platinum is 58·9 old candles per cm². The ratio of the old lumen to the lightwatt can thence be calculated for any given values of C_1 and C_2.

From the relation here obtained, the values of the two ratios are calculated for various values of C_1 and C_2 which have been recommended at various times.

§2. DERIVATION OF AN EXPRESSION FOR THE RATIO OF THE NEW LUMEN TO THE LIGHTWATT AS AN EXPLICIT FUNCTION OF C_1 AND C_2 ALONE

The total luminous flux per unit area of a total radiator, F, expressed in lightwatts cm²., is given in terms of the temperature (°K.), the values of the radiation constants C_1 and C_2, and the relative luminosity function, by the equation

$$F = \pi B = \int_0^\infty V_\lambda J_\lambda' d\lambda, \qquad \ldots\ldots(1)$$

where

$$J_\lambda d\lambda = \frac{C_1 \lambda^{-5} d\lambda}{e^{C_2/\lambda T} - 1}. \qquad \ldots\ldots(2)$$

Here F = total luminous flux per unit area of surface (lightwatts cm.⁻²);

B = brightness (in corresponding units);

$J_\lambda d\lambda$ = total energy flux per unit area from a total radiator, between the wave-lengths $\lambda \pm d\lambda/2$, as given by Planck's equation (watts cm.⁻²);

V_λ = value of the relative luminosity factor at the wave-length λ;

T = temperature (°K.).

Using these equations, and assuming the C.I.E. values for the relative lumi-nosity function (*loc. cit*), values of B for various temperatures between 2000 °K. and 3120 °K. have been calculated and a relation between B and T derived (Caldin, 1945). It is concluded that F can be expressed as a function of T, C_1 and C_2 with con-siderable accuracy by an expression of the form

$$\log_{10} F = a + \log_{10}(C_1/3\cdot703) - (10860 C_2/14330 T), \qquad \ldots\ldots(3)$$

where a is a constant, C_2 is in micron degrees, and C_1 is in erg. sec.$^{-1}$ cm.$^{-2} \times 10^{-5}$. Near the freezing point of platinum, which lies between 2040 °K. and 2050 °K., the value of F is given within ± 0.001 log units (0.2%) by equation (3) with the value $a = 4.775$.

The freezing point of platinum is given on the international temperature scale by the formula (Roeser, Caldwell and Wensel, 1931)

$$\log_e R_1 = \frac{C_2}{\lambda}\left(\frac{1}{T_0} - \frac{1}{T_1}\right), \qquad \ldots\ldots(4)$$

where $T_1 =$ freezing point of platinum;

$T_0 =$ freezing point of gold $= 1336$ °K.;

$R_1 =$ experimentally-determined ratio of the values of J for a given wave-length λ at temperatures T_1 and T_0.

With the value $C_2 = 14320$ micron degrees, the value found for T_1 was 2046.7 ± 1 °K. The estimated error corresponds to an error of $\pm 0.6\%$ in F.

For a small change dC_2 in C_2, equation (4) gives the corresponding change in T_1:—

$$\frac{dT_1}{T_1} = -\frac{dC_2}{C_2}\left(\frac{T_1}{T_0} - 1\right) = -0.53\frac{dC_2}{C_2}. \qquad \ldots\ldots(5)$$

Thus we obtain the following expression for T_1 in terms of C_2:

$$T_1 = 2046.7[1 - 0.53(C_2 - (14320/14320]. \qquad \ldots\ldots(6)$$

Substituting this value for T in equation (3), we obtain finally the following expression for F_1, the total luminous flux per unit area, in lightwatts per cm^2, of a total radiator at the freezing point of platinum:—

$$\log_{10} F_1 = 4.775 + \log_{10}(C_1/3.703)$$
$$- [3.7027 \times 10^{-4}C_2]/[1 - 3.7 \times 10^{-5}(C_2 - 14320)]. \qquad \ldots\ldots(7)$$

Since the value of F_1 at this temperature is 60π new lumens, the ratio M_1 of the new lumen to the lightwatt is given by

$$\log_{10} M_1 = \log_{10} 60\pi - \log_{10} F_1$$
$$= -2.500 - \log_{10}(C_1/3.703)$$
$$+ [3.7027 \times 10^{-4}C_2]/[1 - 3.7 \times 10^{-5}(C_2 - 14320)]. \qquad \ldots\ldots(8)$$

From this equation, so long as the C.I.E. relative luminosity data are accepted, the ratio M_1 can be calculated for any given values of C_1 and C_2. The limits of error are obtained by summing the estimated errors in the use of equation (3) and in the experimental determination of T_1; the result is $\pm 0.8\%$. Assuming in addition the experimental value 58.9 old candles $= 60$ new candles, the ratio M_2 of the old international lumen to the lightwatt is given by a similar equation in which the first term is not (-2.500) but (-2.508).

The ratio M_1 could of course be directly computed for each pair of values of C_1 and C_2, using equations (1), (2) and (4), but such a computation is laborious; the purpose of the present treatment is to show how the previous computations of B (Caldin, 1945) can be used to give a rapid method of computation of sufficient accuracy.

§3. RESULTS FOR CERTAIN VALUES OF C_1 AND C_2

The values of the two ratios, M_1 and M_2, have been calculated for three pairs
of values of C_1 and C_2 which have been recommended at various times.

The results are given in the table below. The values of M_1 should be correct
to about ± 5 units; this uncertainty is mainly due to the experimental errors in
the determination of T_1, the freezing point of platinum. The values of T_1
corresponding to the various values of C_2 are also given in the table.

Authority	C_2 (micron deg.)	C_1 (erg. sec.$^{-1}$ cm.$^{-2}$) $\times 10^{-5}$	T_1 (° K.)	Ratio M_1 (new lumen/ lightwatt)	Ratio M_2 (old lumen/ lightwatt)
International Critical Tables	14330	3·703	2045·9±1	643	631
Wensel (1939)	14360	3·732	2043·7±1	664	652
Birge (1941)	14384·8	3·7403	2041·8±1	683	671

§4. ACKNOWLEDGMENT

The author's thanks are due to Mr. J. S. Preston, of the National Physical
Laboratory, for his suggestions.

REFERENCES

BIRGE, R. T., 1941. *Rev. Mod. Phys.* 13, 233.
CALDIN, E. F., 1945. *Proc. Phys. Soc.* 57, 440.
C.I.E., 1924. *Compte Rendu* (Geneva), p. 67.
JUDD, D. B., 1931. *Bur. Stand. J. Res., Wash.*, 6, 465.
ROESER, W. F., CALDWELL, F. R. and WENSEL, H. T., 1931. *Bur. Stand. J. Res., Wash.*,
    6, 1119.
WENSEL, H. T., 1939. *Bur. Stand. J. Res., Wash.*, 22, 375.

REVIEWS OF BOOKS

Radium Therapy. Its Physical Aspects, by C. W. WILSON. Pp. xi+224.
    (London : Chapman and Hall, Ltd., 1945.) 18s.

Dr. Wilson has attempted with considerable success to fill a gap of long standing in
the literature of radiation physics applied to medicine. The publication is also timely,
as the author points out in his preface, in view of the expected implementation in the near
future of the Cancer Act (1939), when the wider application of radiotherapy is likely to
attract more workers, medical and otherwise, to this interesting field.

The book should amply meet the needs, as regards radium therapy physics, of the
postgraduate student of radiology, particularly, of course, if he intends to specialize in
radium therapy. It provides at the same time a handy reference volume and data book
for the practising radium therapist and the hospital physicist, however familiar they may
be with much of its contents. It is written with the clarity of style and the close attention
to detail which has characterized all Wilson's many publications on the subject. Little
work of importance in the last decade or so has been overlooked, and the book abounds
in references to the literature.

The scope of the book is clear from the chapter headings : (1) Radium : its properties and application to medicine ; (2) The interaction of high-voltage radiation and matter ; (3) Gamma-ray dosimetry : methods and theory ; (4) Gamma-ray dosimetry : apparatus for dosage measurement ; (5) Radium therapy with surface applicators (plaques) ; (6) Cavity and interstitial radium therapy ; (7) Radium teletherapy ; (8) Protection. There are four appendices, valuable as aids to practical dosage calculation.

While the general treatment is on very practical lines, theoretical aspects are by no means neglected. Thus, while details of the construction and design of " air-wall " ionization chambers are fully described, recent work on " air-wall " materials and chamber wall-thickness is clearly presented from the theoretical standpoint. The chapter on radium teletherapy is particularly good, and here the author has drawn considerably on his wide and intimate association with the many physical problems involved in the technique.

It is not surprising to find, in the chapters dealing with surface radium applicators and interstitial gamma-ray therapy, that the well-known dosage systems of Paterson and Parker receive much attention and are described in detail. The dosage graphs are given, but, as in the original published work, they are too small for practical use. In this connection a useful addition, in appendix form, would be the table published by Oddie in 1941, relating radium-skin distance, area and milligramme-hour dose for surface applicators, preferably recalculated to 0·5 mm. Pt screenage (instead of 0·8) and adjusted to 8·3 r. per hour (instead of 8·0) for the value of the dosage rate at 1 cm. from a point source of 1 mg. Ra element with this screenage.

The layout of the book is excellent and the line diagrams are numerous and clear. The production would be considered satisfactory even on peace-time standards, and the price is very reasonable. In future editions, when the paper shortage is less acute, the addition of some blank pages for notes alternating with squared paper would be useful at the end of the book.

It is abundantly clear that the author, who is Physicist in the Department of X-ray and Radium Therapy at Westminster Hospital, fully appreciates the clinical problems which confront the radium therapist. The fruitful co-operation of physicists and others engaged in radiation therapy depends largely upon a mutual appreciation of each other's problems and difficulties. The author expresses the hope that his book will help to make the physicist's work appear as a more co-ordinated whole than hitherto. There is little doubt that it will do so.

D. E. A. J.

Fluid Dynamics, by R. VON MISES and K. O. FRIEDRICHS, with supplementary notes by S. BERGMANN. (Rhode Island : Brown University.)

There is some doubt about the propriety of reviewing these papers, as copies can only be obtained by writing to Brown University. The papers consist of " Roneoed " notes of lectures given at courses of advanced instruction and research in mechanics at Brown University in 1941 and 1942. They are sufficiently comprehensive to be published as a book, but there is no indication that this is intended. This would be useful, as the notes provide a good exposition of the field covered and would help the student attending the lectures.

Fluid dynamics is treated from a strongly theoretical standpoint, as might be expected from the reputation of the authors. The first chapter gives a résumé of the standard theory of the motion of an ideal fluid ; vector methods are used throughout and the treatment is thorough yet concise. This chapter is followed by a chapter on motion in two dimensions, with special reference to the aerofoil ; this treatment follows the lines developed by von Mises in 1917 in stressing the geometric aspects of the conformal transformations and of the resulting aerofoil section. It is doubtful whether the student should be burdened with this kind of treatment which is hardly ever used in modern developments of aerofoil theory. Chapter 3 gives the basic theory and elementary application for the wing of finite span, and includes a full account of the properties of vortex lines and sheets.

The first three chapters are written by von Mises, and the next two chapters, discussing viscous and compressible flow, are written by Friedrichs. The basic theory of viscous

flow, in particular the validity of the boundary-layer equations, is fully discussed, while the applications of the theory are treated as examples. The treatment of the instability of laminar flow is more thorough than in any available book on aerodynamics and is clearly one of the subjects of special interest to the author. This subject has occupied the attention of mathematicians and physicists since the time of Osborne Reynolds ; due to the complicated equations, it has been difficult to deduce conclusive results, but it is gradually becoming accepted that Tollmien's 1929 investigation and those following from it are valid.

The discussion of compressible flow covers flow in channels, flow across and in shock waves, linearized theory of subsonic and supersonic flow, and the flow round corners.

The concluding section, by Bergmann, on the hodograph method in the theory of compressible fluid flow is more specialized than most of the rest of the " book ". The hodograph method enables the equations of motion to be linearized and solutions in terms of hypergeometric functions to be obtained. In the hands of Karman and Tsien it has yielded general approximate formulae relating the flow of a compressible fluid to the flow of an incompressible fluid round a body. Ringleb and Karman have shown that the limiting speed, above which solutions of the equations cannot be applied, is associated with the occurrence of infinite accelerations in the field ; this occurs only at speeds well above the local speed of sound. The question whether this phenomenon is the cause of the development of shock waves is unlikely to be answered by mathemetical analysis only, as some instability of the steady supersonic flow may be present at lower speeds.

The hodograph method is well adapted to deal with general problems, but its application to determining the flow round definite bodies is difficult, as the body shape tends to vary with the speed of flow unless troublesome corrections are made. The only practical alternative to the hodograph method is numerical analysis, but it is always difficult to deduce general conclusions from numerical data.

In conclusion it is to be hoped that the material contained in the lecture notes will be published in due course as a book.   H. B. SQUIRE.

GEOMAGNETIC SECULAR VARIATIONS AND SURVEYS

By J. A. FLEMING,

Director, Department of Terrestrial Magnetism,
Carnegie Institution of Washington

Third Charles Chree Address, delivered 6 December 1945

ABSTRACT. The secular variations of the Earth's field call for frequent determinations of the geomagnetic elements at many selected stations on land and at sea. So far, general world magnetic surveys have of necessity been restricted to the surface of the Earth. Previous surveys, including those of the Department of Terrestrial Magnetism of the Carnegie Institution of Washington, have led to great improvement in our knowledge of the distribution of the field, especially during the past few decades. During the war, much attention has been devoted to the study and analysis of data well distributed over the Earth's entire surface at more than 10,000 stations. These have resulted, for the first time, in the preparation of accurate isoporic charts, that is, charts of equal annual rate of change for magnetic declination, inclination, and the horizontal, vertical, eastward, northward and total components of the field for the four epochs of 1912·5, 1922·5, 1932·5, and 1942·5. The motions of the maximum and minimum isoporic foci during these four epochs indicate the complexity of the secular changes and interpretations.

Isomagnetic charts based upon surface observations must always be limited, so far as faithful depiction of the field is concerned,-because of the impracticability of obtaining observations at an infinite number of stations and of deductions for both the regular and irregular changes in the field. Progress in instrumentation during the war on the applications of the geomagnetic field of the Earth have produced improvements which make feasible the early realization of magnetic surveys by airplanes at several different levels proposed by Professor Sydney Chapman in the first Charles Chree Address. Some of the potentialities, possibilities, and needs for intense national and international coordination in magnetic surveys by plane are reviewed.

§1. INTRODUCTION

I AM humbly grateful to your Society for making me the third recipient of the Charles Chree Medal and Prize. The award is, I take it, largely a recognition of the geomagnetic researches which my colleagues and I have made at the Department of Terrestrial Magnetism during more than four decades with the generous support of the Carnegie Institution of Washington. The two previous awards were to scientific colleagues and friends (see Chapman, 1941, and Schonland, 1943), whose work has constantly inspired investigators everywhere, and especially those of us in the United States who have been associated with the Department of Terrestrial Magnetism. The award affords further evidence of that scientific co-operation of British and American physicists and geophysicists which has been so vital in forwarding our common war effort, particularly in geomagnetic applications.

Progress in geomagnetism owes much to Charles Chree as Superintendent of the Magnetic and Meteorological Observatory at Kew from 1893 to 1925, and thereafter until his death in 1928. He and his contemporaries—Schmidt, Bauer, Van Bemmelen, and Moos—built well on foundations laid by their eminent predecessors in the task of portraying the complex phenomena of the magnetic and electric fields of the Earth through surveys on land and séa, investigations in the laboratory, and theoretical hypotheses and interpretation. As a colleague and friend, Chree influenced greatly the work of Bauer in 1899 through the expansion of geomagnetic research in, and survey of, the United States Coast and Geodetic Survey, and again in 1904 through the realization of Bauer's plan for a Department of Research in Terrestrial Magnetism submitted to the Trustees of the Carnegie Institution of Washington. His wise counsel carried great weight and played a large part in the subsequent development and growth of the Department. Chree's share in international and national progress in our field is well known to all of you. His commonsense treatment of factual information gathered in all parts of the world—especially in polar regions—was outstanding; it has contributed much to applications of geomagnetic research touching so many lines of human peace-time and war-time endeavor.

The foundation and provision by Miss Jessie Chree of the Charles Chree Medal and Prize through the Physical Society not only serve to honour the memory of Chree's accomplishments but also to stimulate future workers and progress in our fields.

The first Charles Chree Address, in 1941, by Professor Sydney Chapman gave a scholarly general account of the magnetic field of the Earth. I have chosen for my subject *Geomagnetic secular variations and surveys*, and will attempt to review at greater length world magnetic surveys and potentialities so prophetically outlined in Part III of the first Address.

§2. GEOMAGNETIC SECULAR VARIATIONS

Several centuries of speculation and research on the old and challenging problem of the origin of the Earth's magnetism and its changes have yielded as yet no adequate explanation. Much has been learned from past attempts at explanations, and we have also added, in the course of time, much new and rich material relating to the description of the phenomenon. This material has illuminated greatly the picture of what we are trying to explain, but understanding of the cause of geomagnetism is emerging relatively slowly.

In spite of vast improvements in our knowledge of the geomagnetic field, that knowledge remains insufficient for various theoretical and practical problems. For example, Schrödinger (1943) has recently attempted to use the data from geomagnetism to test some theories of physical fields which he has developed. For these tests he required information on the magnitude of the external field and of the non-potential field. It was important to know the magnitudes of these quantities to a moderate degree of precision, but those who are working in geomagnetic research recognize that we cannot even be sure, on the basis of existing observations, that there is an external field or a non-potential field.

The magnetic field varies with time (Fleming, 1938a; Barker and McNish, 1938). Its delineation at any instant is hence difficult because it is not practical

to undertake its simultaneous measurement at all points of the Earth's surface. There are some 70 magnetic observatories (figures 1 and 2) where continuous simultaneous measurements of the field are made, but the complexity of the field and its changes with time do not permit satisfactory interpolation of values in intervening regions. The records at observatories must be supplemented by observations at many isolated points. The density of observations, however, per given area of survey is uneven, and great oceanic areas are sparsely represented. Measurements at these stations on land or sea are usually made on one day only, and the resulting data form the basis for isomagnetic charts.
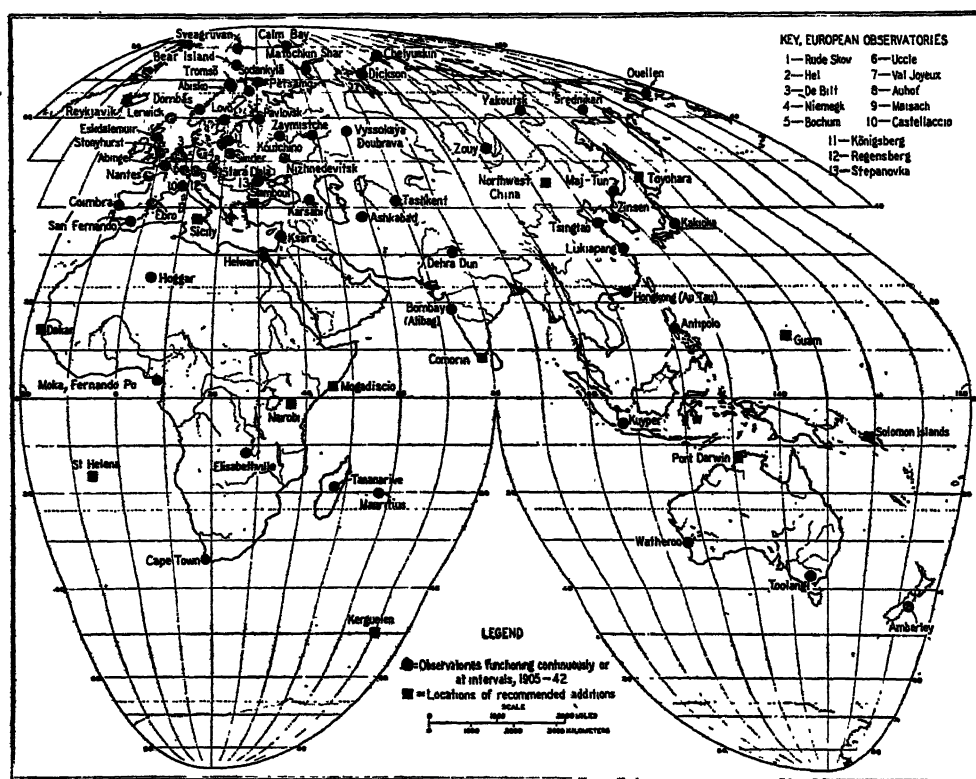
Figure 1. Geomagnetic observations, Eastern Hemisphere, functioning continuously or at intervals, 1905–42, and recommended additional locations.

The world magnetic survey of the Carnegie Institution of Washington from 1905 to the beginning of the war improved knowledge of the general field and its changes, which are slow when compared with the life-span of a human being but are very rapid when compared with astronomical or geological scales of time. Nor are these secular changes small. During the past half-century a gradual change amounting to 10% of the maximum value of the Earth's field has occurred off the Guinea coast. By way of further illustration, consider a few indications based on long series of observations. The actually observed changes in magnetic direction at London and at several other places as determined from earliest known measurements, including those for 1945, are shown in figure 3. The figure shows also

notable changes in the position of the magnetic equator since 1700. The progression in direction of the compass by centuries is non-uniform and notably large. The general aperiodic character of secular change in direction is confirmed by the measured directions of magnetization of particles locked in annual varved deposits in lakes and age-old deposits in oceans.

Radical changes in the estimated position of the lines of zero magnetic declination (agonics) are noted since the year 1500. The line along which the compass
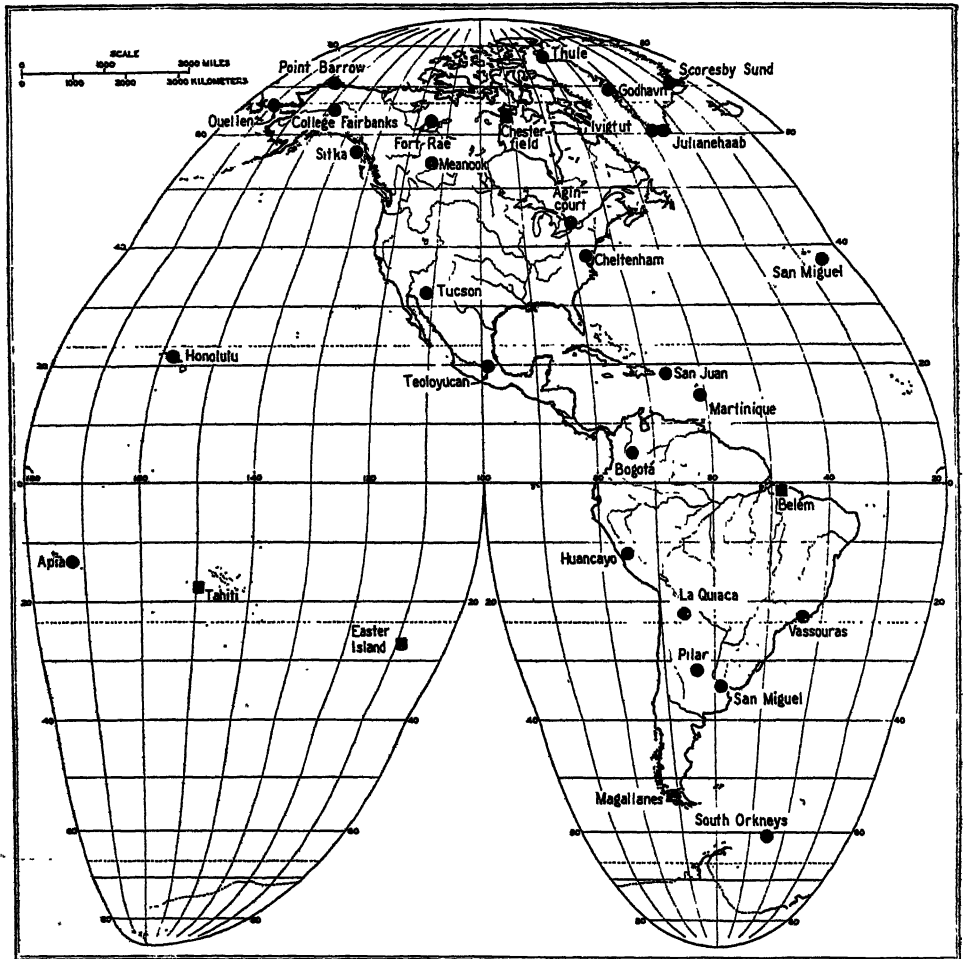
Figure 2. Geomagnetic observations, Western Hemisphere, functioning continuously or at intervals, 1905–42, and recommended additional locations.

points true north (figure 4) has shifted westwards from Suez to Chicago in 445 years, only a short interval in the Earth's history.

Clearly, secular change involves the general magnetic field of the Earth, and any attempt to explain either phenomenon must also account for the other.

There are two hopeful present approaches. One is the derivation of more adequate and complete data on the observed large-scale changes in the Earth's magnetism. The other is the improved description of the permanent field itself.
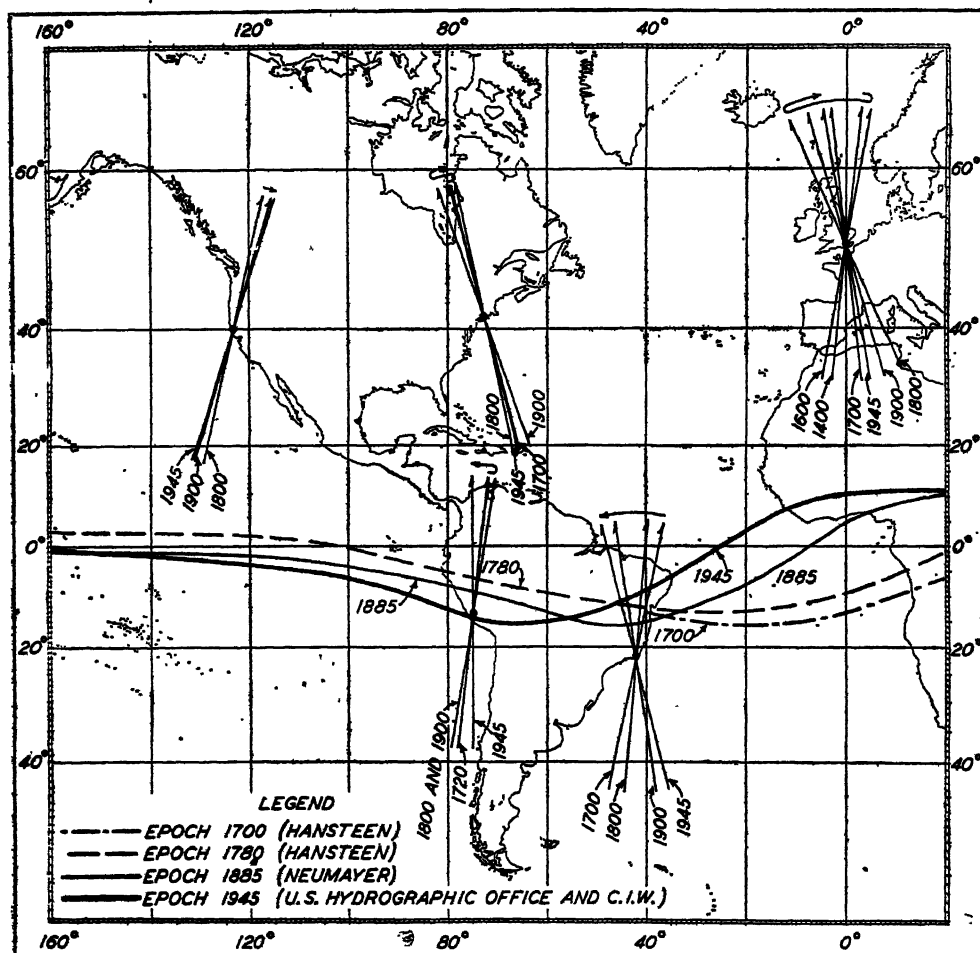
Figure 3. Secular variation of magnetic equator in Western Hemisphere from 1700 to 1945 and of magnetic declination at London, Eastport, San Francisco, Callao, and Rio de Janeiro, from earliest observed values.
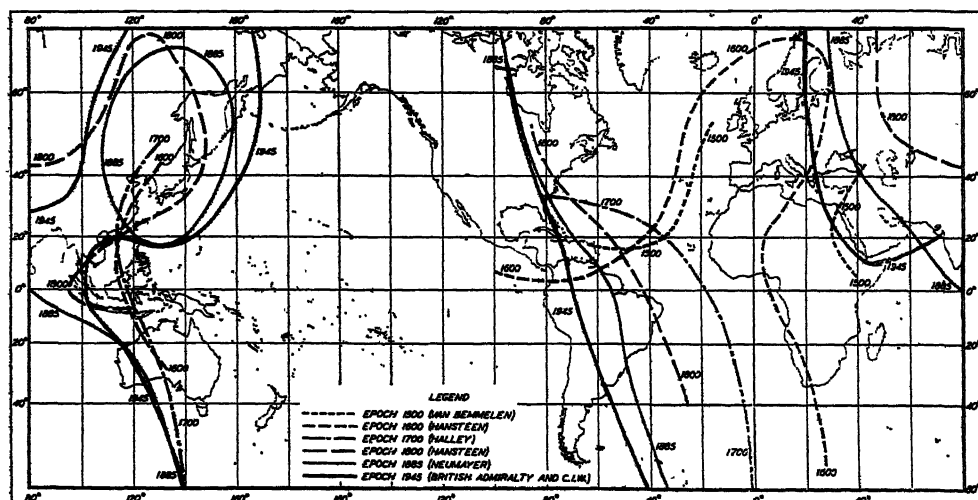
Figure 4. Progressive change (secular variation) of lines of zero magnetic declination (agonics). (The dots indicate the side of the agonic lines on which the declinations were easterly.)

The special type of the research required calls in a unique degree for world-wide coordination of data and experiments. No single, well-planned experiment or observation can supply the solution. Observations must be made in all parts of the world and continued for a long period of time. Techniques for the organization and interpretation of the data must be developed and experimental research must be conducted along lines which will supply information on basic physical problems related to this subject.

In the reduction of magnetic observations to epoch, the aim is to determine at all stations (at epochs not too far removed from those at previous times of observation) the intensity and direction of the Earth's field freed from the effects of small superposed fluctuations in intensity. For practical reasons the problem of describing the Earth's surface field is simplified by mapping only the main or permanent field. The strength of the main field at any time is called its *normal value* at that time.

Although the rough and general description of the geomagnetic field in terms of its normal value as shown on magnetic charts is comparatively easy, the derivation of the normal values themselves is complex and difficult. To obtain these normal values it is first necessary to remove from the individual magnetic observations the contributions of extraneous fluctuations not a part of the secular variation. Often the extraneous fluctuations are large enough to affect seriously the estimates of secular change obtained from two or more observations made in different years at the same place. These fluctuations may thus make difficult or impossible a reliable estimate of the normal values (since the secular change may not be determined accurately either with respect to magnitude or sign) at epochs other than those of observation. When the secular change has been corrected for fluctuations in field, so that it is known with accuracy, the magnetic observations are then readily reduced to the epoch desired for a magnetic chart.

Of fundamental importance in charting the Earth's magnetic field for a given epoch are charts showing rates of annual secular change, that is, isoporic charts. Too little attention has been given by various organizations responsible for the preparation of world isomagnetic maps to the importance of constructing isoporic charts. This importance arises because it provides the only feasible means of enhancing and extending the value measured at a station, say in 1920, to determine the value for the station at a later time, say 1945.

The first comprehensive world isoporic charts were prepared by Fisk (1929 and 1932) and by Fisk and Fleming (1931) at the Department of Terrestrial Magnetism for the epoch 1922·5 (specimen given in figure 5). These give the average annual secular change in the magnetic elements and their components, namely, declination, inclination, and the horizontal, vertical, total, northward, and eastward components of the field.

Fisk's isoporic charts, which have been widely used, were naturally tentative in a number of regions. They did not take into account corrections of the observed data for geomagnetic fluctuations. They are also in several respects mutually inconsistent with the known nature of the geomagnetic field; and the singularities in field near the poles could not be taken into account because of the sparsity of data in high latitudes.

The danger of using these charts today in mapping is that the pattern of the

isopors changes rapidly. For instance, in declination there is scarcely the slightest resemblance between the isopors of 1922 and those of 1942 in many parts of the world. Uncertainties of this kind account for relatively large discrepancies on some charts to actually observed values. In other words, it is necessary to know

Figure 5. Isoporic chart for magnetic declination in minutes per year, approximate epoch 1922. Prepared at the Department of Terrestrial Magnetism, Carnegie Institution of Washington, by H. W. Fisk, 1930 (slightly revised, October 1942).

not only the pattern of the isopors at a given time but also the trend of change in pattern. This trend can only be derived by mapping the isopors for several different epochs. This in turn requires that there be maintained adequate surveys at regular intervals and continuous observations at magnetic observatories.

§3. GEOMAGNETIC SURVEYS

The accuracy of magnetic observations depends mainly on the type and character of equipment, plan and technique of operation, and skill and care of the observer. Thus measurements of a half-century ago are less accurate than those made in more recent years because of improved and superior instruments and techniques. An important illustration of this trend was the introduction of the earth-inductor for measuring magnetic dip to supplant the dip-circles of earlier generations, with substantial increase in accuracy. Loss of accuracy may result from deterioration of an instrument after years of use. The observational programme and the number of measurements required at a station vary with each data-gathering organization. Although procedures of measurements are well standardized on the whole, the results obtained by different observers show small differences when using the same instrument.

There are many sources of error: for example, magnetic impurities in the metal parts of an instrument; changes with time in adopted standards, such as change in the moment of inertia of the oscillating magnet-system in measuring

horizontal intensity; changes with time in deflection distances; wear in bearings and brushes; unspecified changes in instrumental constants not readily ascertained but detected by inter-comparisons with carefully maintained standard instruments; poor levelling and influences of temperature-changes during measurements; infrequent inter-comparisons of magnetic instruments with reliable standards; and faulty computations of results—the last often very serious. Fortunately, in most surveys, errors from these sources are of minor character and less important than the influences of extraneous fluctuations on data used in mapping the main field and its secular vaiation.

The large portion of the Earth's surface covered by the oceans makes the determination of accurate values at sea a major objective of the world-wide magnetic survey. It was not until 1905 that full realization of this objective had a definite beginning through systematic oceanic surveys then sponsored by the Carnegie Institution of Washington.

As you are well aware, the first attempt to accomplish a magnetic survey at sea was the expedition of Halley between 1698 and 1700 (Chapman and Bartels, 1940; Fleming, 1940). He made several voyages in the North and South Atlantic oceans, determining magnetic declination only. The results were embodied in his chart "Lines of equal magnetic variation" of the Atlantic for the year 1700—the first isomagnetic chart. In the past 100 years extensive magnetic surveys have been made by many countries. Important undertakings following were in the *Erebus*, the *Terror*, and the *Pagoda* between 1840 and 1845, chiefly in southern waters; on these, all elements were observed utilizing the Fox dip-circle for measuring magnetic inclination and intensity. The frigate *Novara* obtained magnetic declinations while circumnavigating the globe in 1857–60. Observations of the three magnetic elements were made over various oceans during the notable cruises of the *Challenger* in 1872–76 and of the *Gazelle* in 1874–76. More recently, observations were also made by the naval services of the various countries and by antarctic expeditions, notably the *Discovery* and the *Gauss*. The accompanying charts (figures 6 *a*, 6 *b*, 6 *c*) show the tracks of the chief vessels on which magnetic observations were made from 1839 to 1916.

All these observations were of varying degrees of accuracy set by available instruments and by the disturbing factors originating in the magnetic character of the vessels. The distribution, both as regards position and epoch, was not such as to yield coordinated charts applying to definite epochs. Therefore, in planning in 1904 to fill in the gaps in data for the remaining regions on land and sea for the world magnetic survey, the Carnegie Institution of Washington gave careful consideration to the oceanic survey. The Institution's earliest work was with the chartered vessel *Galilee* in the Pacific Ocean during 1905 to 1908. The *Carnegie* was designed in 1908 primarily for magnetic surveys and investigations, and the vessel and equipment were completed in 1909. Before the loss of the *Carnegie* at Apia in November, 1929, the ten cruises of that vessel and the *Galilee* had traversed some 362,000 nautical miles and from tens of thousands of individual observations had obtained values of declination at over 3800 stations and of both inclination and horizontal intensity at over 2300 stations. Figure 7 *a* shows in a general way the scope of these activities from 1905, whilst figure 7 *b* shows a chart of horizontal intensity obtained from the existing sources of information in 1940.

On the side of practical application, the increasing use of the oceans in the commerce of nations by sea and air makes the continuation of the survey a matter of international concern and benefit. First among those investigations demanding

Carnegie, 1911–1916 —•——•——•
Novara, 1857–1860 •••••••••••••
Gauss, 1902–1903 ——•——•——•——
Erebus and Terror, 1839–1843 ••——••——••——••
Challenger, 1872–1876 —————————————
Discovery, 1902–1904 ••••••••••••••••
Pagoda, 1845 ——————•——•——
Gazelle, 1874–1876 •——•——•——•——
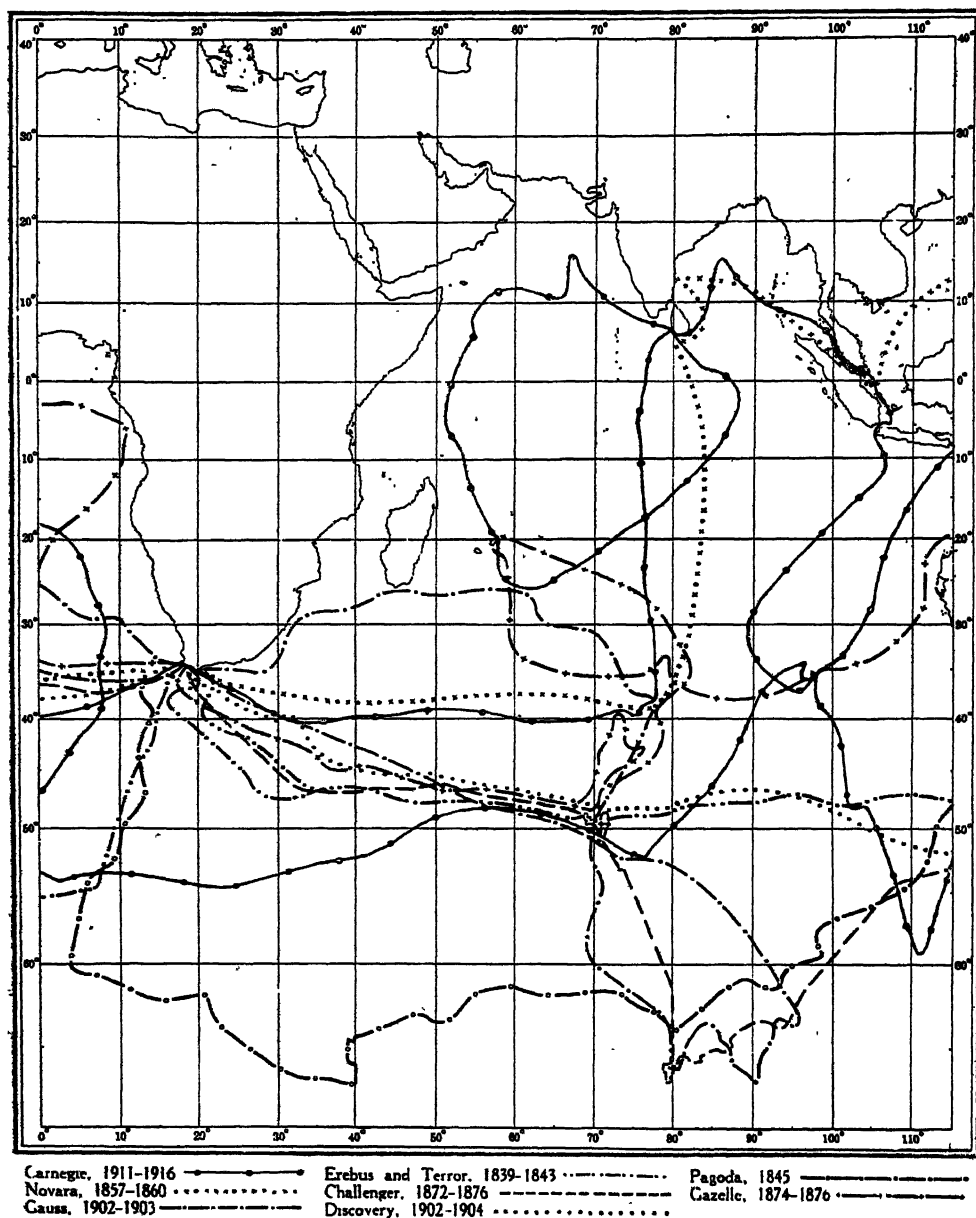
Figure 6 *a*. Tracks of chief vessels on which magnetic observations were made in the Indian Ocean, 1839–1916.

continuation of the surface magnetic survey is the determination of the acceleration of secular variations; accumulated data show that they cannot be extrapolated reliably over periods as long as five years. A definite control is necessary for a number of epochs to facilitate the investigation of causes producing and governing
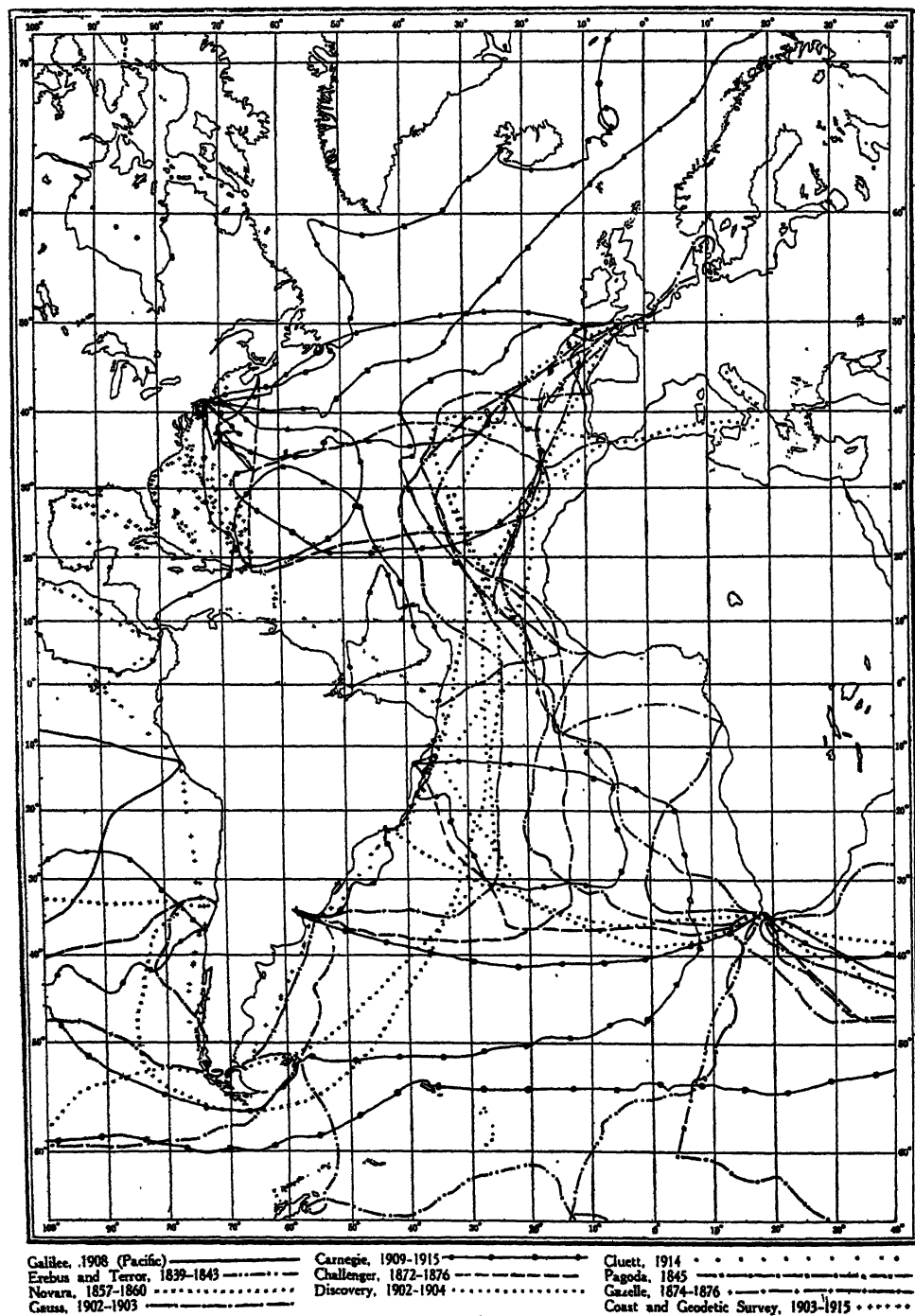
Figure 6 *b*.   Tracks of chief vessels on which magnetic observations were made in the
Atlantic Ocean, 1839–1916.

these progressive changes, which, it appears, would be favoured by accurate knowledge of their accelerations and distribution. The observations at repeat-stations on land have helped greatly, in conjunction with more detailed results of magnetic observatories, in delineating the complex patterns of secular change on

Galilee, 1905–1908	Carnegie, 1911–1916	Cluett, 1914 (Atlantic)
Erebus and Terror, 1839–1843	Novara, 1857–1860	Challenger, 1872–1876
Gazelle, 1874–1876	Discovery, 1902–1904	Coast and Geodetic Survey, 1903–1915
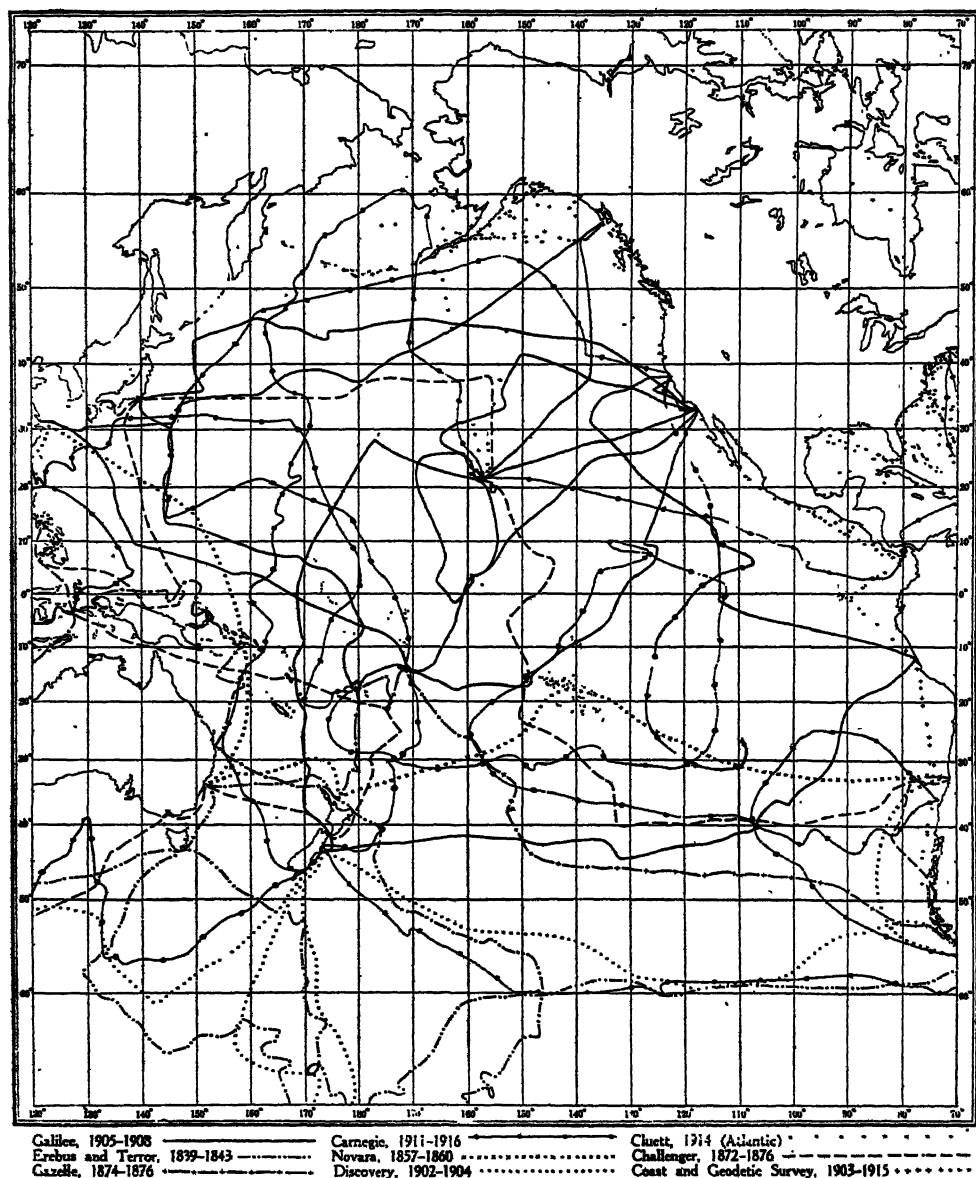
Figure 6 c.   Tracks of chief vessels on which magnetic observations were made in the Pacific Ocean, 1839–1916.

land.   By continuity these have served in some uncertain measure to bridge the gap across the oceans as well, when supported by the secular changes given by the coarser and less accurate ocean observations. It should be noted here that, though the cruises of the *Galilee* and *Carnegie* afforded usually only rough indications of

secular change in field, there were discovered great discrepancies in existing charts of the field itself.   Great gaps in information remain in polar regions; for example,
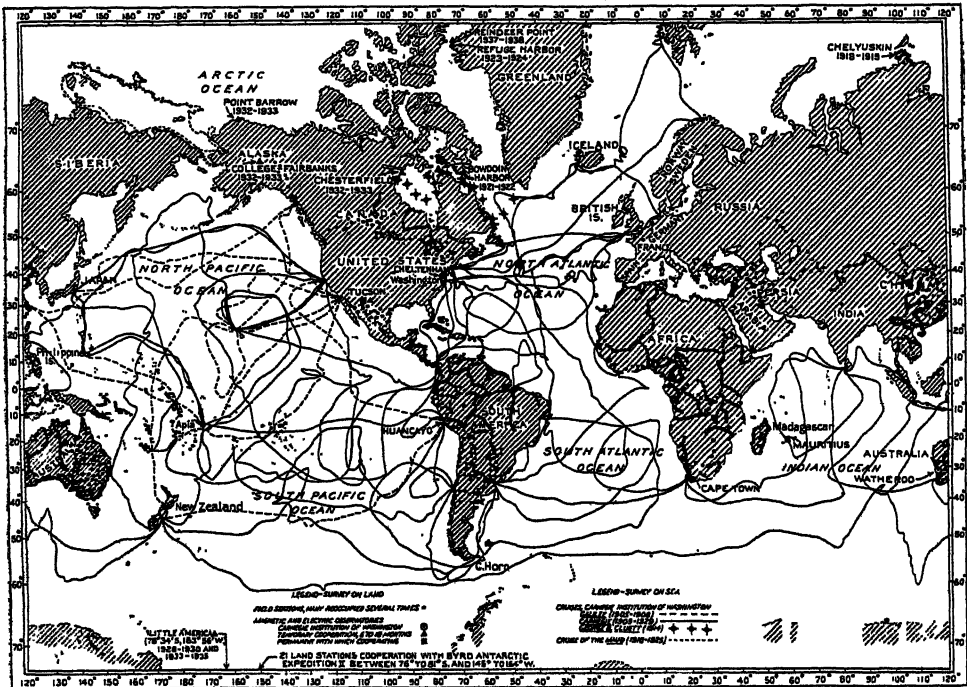
Figure 7 *a*.   Magnetic survey operations of Department of Terrestrial Magnetism, Carnegie Institution of Washington, from 1905.
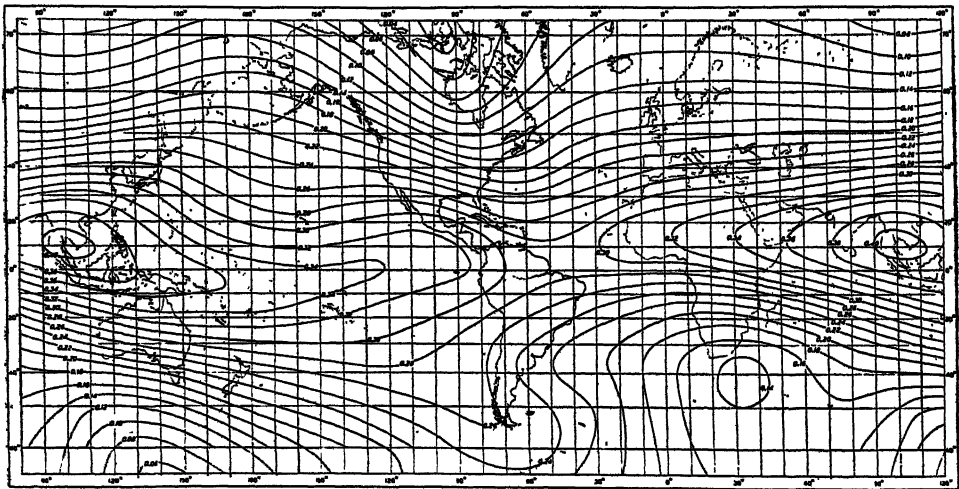
Figure 7 *b*.   Isomagnetic chart of horizontal intensity of the earth's magnetic force for the year 1940 (expressed in centimetre-gram-second units).

the south polar regions have only about four badly distributed stations yielding inferior indication of secular change since 1900.

It is gratifying that the British Admiralty designed and constructed the non-

magnetic vessel *Research* to continue the surface survey at sea and to fill in those serious gaps in the present data caused by the rapid change in the secular variation in certain regions of the Earth since the earlier surveys.   But the task in the geophysical survey of the oceans is so great that other hydrographic services of maritime nations should also provide suitable surface or air vessels with equipment and personnel to take appropriate and coordinated share in the world survey.

§4. ISOPORIC CHARTS AND GEOLOGIC FEATURES

That irregularities of the distribution of the Earth's magnetic field over continental areas may bear definite relation to geologic features of the Earth's crust has been recognized for a long time.   It is only in comparatively recent years that such magnetic data have been intensively considered as an aid in the determination of the character of geologic substructure.   From accumulated observations it is possible to account for a considerable part of the Earth's main field on the basis of a uniformly magnetized sphere.   The difference between such a uniform or normal magnetization and that actually existing serves to bring out the general irregularities as a residual field.   The Earth's outer crust or shell, perhaps 25 or more miles thick, is not homogeneous and, therefore, is not uniform in its magnetic behaviour. Thus there are many regions of local magnetic disturbance, such as are caused by magnetic ore-deposits, some so great as to give to local poles and other irregularities of intensity twofold or even threefold the normal value in such regions.   Anomalies in Russia, in Germany, and in New South Wales are a few striking examples of such magnetic irregularities.   The great anomaly in South Africa has been apparently retained for over 100 million years by the huge Pilansberg system of Palæozoic volcanic dykes (figure 8); fourteen vertical dykes are found to be strongly magnetized in a direction opposite to the present-day magnetic field of the Earth (Piggot, 1938).   In this general region of South Africa, the magnetic horizontal component of the Earth's field has decreased 16% of its present value in a 30-year period !

The anomalous magnetic features thus observed over continental areas must have their counterparts in the great oceanic basins (Fleming, 1938 b).   We may therefore look confidently to such features as an aid in the study of oceanic structure.   Indeed, observations made on isolated islands and island groups almost invariably indicate abnormal magnetic conditions, and the magnetic survey of the oceans has revealed such conditions at sea.

However, the application of magnetic-survey results to the investigation of crustal features of the ocean bottom depends upon the existence of a difference in substructural physical properties as well as upon the magnitude of this difference. Here enters the magnetic susceptibility of rocks, some of which are strongly magnetic, as iron-ores and igneous and metamorphic rocks, and some weakly magnetic, as most sedimentary rocks.   Here, then, one must think of the observed magnetic anomalies as a kind of general reconnaissance method to indicate areas for other more intensive geophysical studies.   However, the magnetic method of approach lacks depth-control, and, therefore, the size of geologic features must increase in proportion to depth if we are to detect such features.   Thus, unless the magnetic anomaly observed at sea is over shallow water or, if over deep water, is of enormous proportions, there may not be great hope of immediately deriving exact

information regarding particular areas of bottom substructure from distribution
data as shown only by isomagnetic maps.

  The secular-variation change of the Earth's magnetic field is a phenomenon
which has large potential value in the study of the Earth's crust.   Herein is the
great need of continuance of the accurate world surface magnetic survey, that we
may determine the character of such changes with epoch and their accelerations at
various places.

  Ever since the discovery of the change in the Earth's magnetic field through the
centuries, investigators have noted the apparent symmetry and regularity of
secular variation when attention is confined to a single station or to a small part of
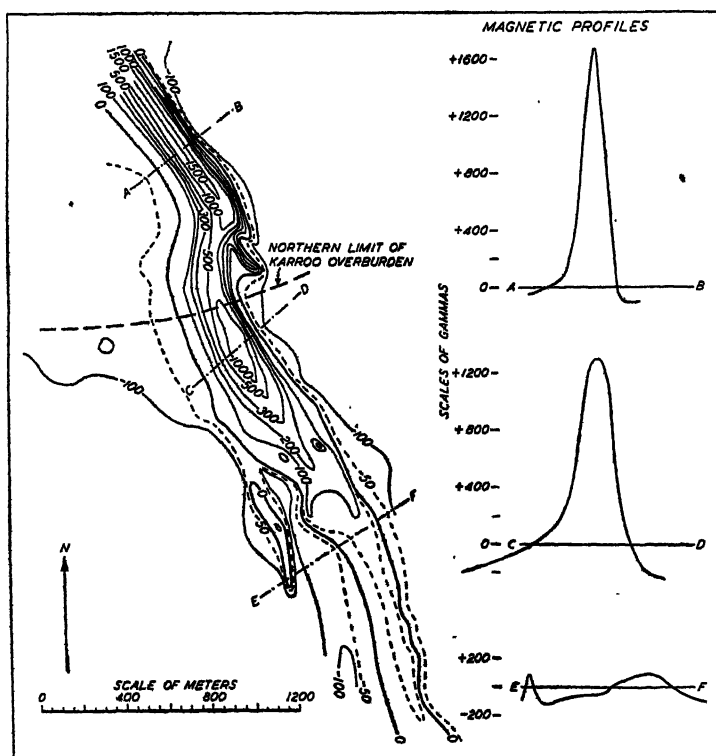
Figure 8.   Magnetic anomaly of Pilansberg Dyke System, southern Transvaal, showing
           anomaly of vertical intensity expressed in gammas (after Gelletich).

the Earth's surface such as western Europe—only one-hundredth of the Earth's
surface.   Considering, however, the actual observational data when the whole
Earth is taken into account, secular variation is not so simple a phenomenon that
it may be explained by periodic movements of the Earth's magnetic poles and the
like, but is a regional phenomenon connected with a large geological structure—
oceans and continents—of the deeper layers of the Earth's crust.   .

  The foci of rapid isoporic changes raise questions of broad geophysical
significance. A large part of these changes may have origin in influences impressed
by forces ceaselessly at work within the Earth.   The rise and fall of the rate of
secular change and the slow expansion and then the gradual retraction of the areas
within which there have been excessive alterations in the magnetic elements are

significant of such changes. The distribution of isoporic foci (figures 9 a, 9 b) is practically all in the hemisphere containing the great land-masses with the intervening Atlantic Ocean. Such foci as are found in the Pacific Ocean are of moderate intensity and not well defined. These relations to the surface structure of the Earth can be scarcely accidental; it is natural to suspect that there is a causal relation between crustal or subcrustal movements and conditions and these magnetic manifestations.
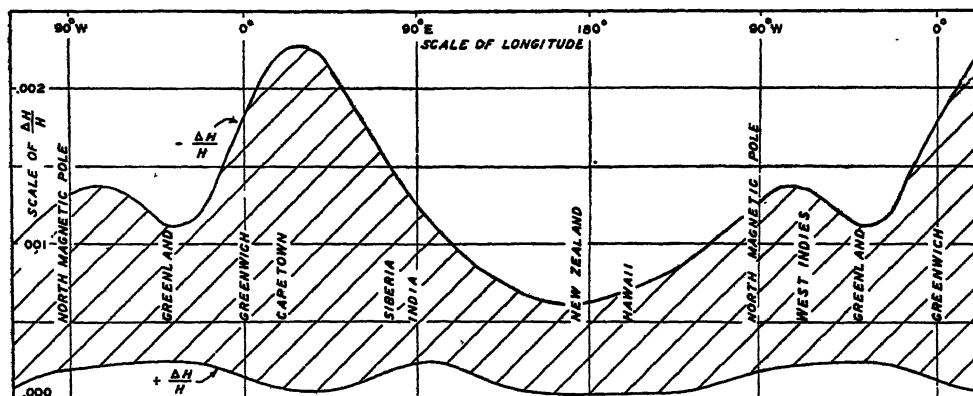
Figure 9 a. Longitudinal distribution of proportion of annual change ($\Delta H/H$) of horizontal intensity (after Fisk). (Lower curve represents average positive values of $\Delta H/H$ in each lune between meridians, while upper curve represents numerical magnitude of average negative values ; thus shaded areas between curves are measures of excess of negative over positive annual change.)
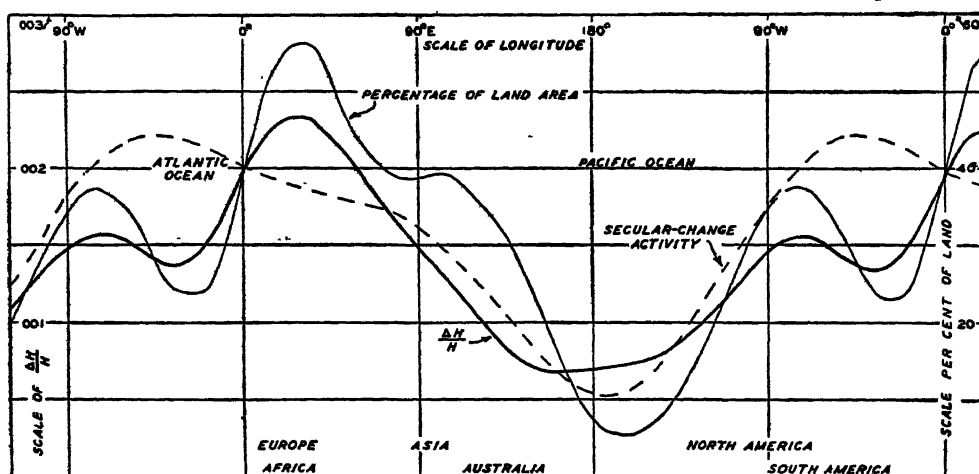
Figure 9 b. Variation with longitude of $\Delta H/H$ (annual change averaged without regard to sign), of distribution of proportion of land and water areas, and of secular-variation activity approximately determined by density of distribution of isoporic lines (after Fisk).

As an example, attention may be called to the apparent diminution of the intensity of the Earth's magnetic field which is marked over oceanic areas, especially in the Southern Hemisphere (Fisk, 1932). The interpretation of such data must be important in geophysical and geological research to advance understanding of Earth phenomena.

J. A. Fleming

Thus the observed longitudinal distribution of magnetic secular variation agrees with that of land areas—as witness the moderate rates of annual change over the Pacific as compared with those over the Atlantic and adjoining continental areas.   A definite control is necessary for a number of epochs to facilitate investigation of causes producing and governing these progressive changes.

Further data bearing on correlation between the surface-distribution of secular-change activity promise conclusions concerning isoporic processes localized in the crustal layer.   Thus continued secular-variation surveys at sea should bring together seismic and magnetic methods of approach to crustal adjustments and possibly gravimetric work.

Secular-variation information may be extended to remote times for which there are no observations.   When lava cools and solidifies, following a volcanic outburst, it takes up a permanent magnetization dependent upon the orientation of the Earth's magnetic field at the time.   This, because of small capacity for magnetization in the Earth's magnetic field after solidifying, may remain practically
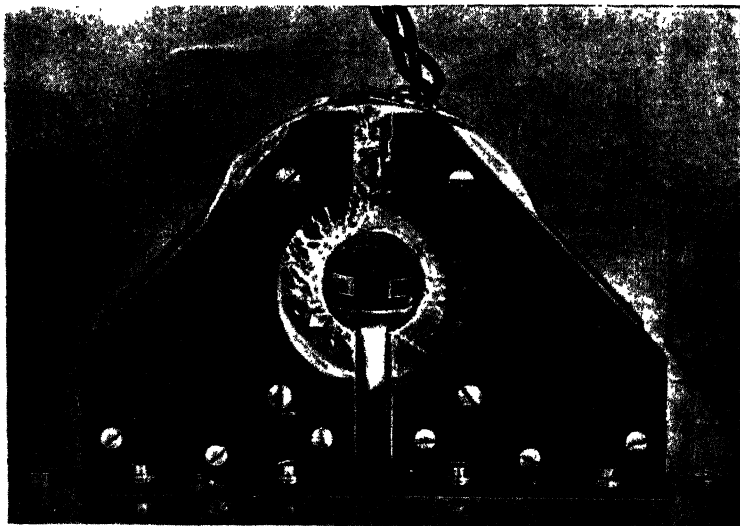
Figure 10.   Apparatus for measuring magnetic moments and orientations of samples of varves and of ocean-bottom cores.

constant.   Thus the direction of the originally acquired permanent magnetization can be determined by laboratory tests provided every detail of the orientation of the mass tested is carefully noted and marked when it is removed.

Just as lavas cooling through the Curie point in the presence of a magnetic field become magnetized in the direction of the field, so sediments containing microscopic magnetic particles also assume the direction of the prevailing magnetic field. This is truly "fossil" magnetization, for the magnetic field of the epoch of deposition is embedded in the sediment as surely as the organic remains of that epoch. Whether this fossil magnetism remains as a connate property of the rock or whether it is destroyed by subsequent events, even as organic fossils are sometimes destroyed, depends upon the post-natal experience of the sediment.   In so far as it remains, it may supply an important record of the Earth's magnetism and an important clue to the history of the rock.

The measurement of the Earth's magnetism at great depths in the sea may now be consummated by magnetic examination of the core-samples taken with Dr. Piggot's apparatus described in 1938. In order that the fossil magnetization of the specimens may not be disturbed, a method of measurement has been developed for tests of high sensitivity and accuracy without submitting the specimens to
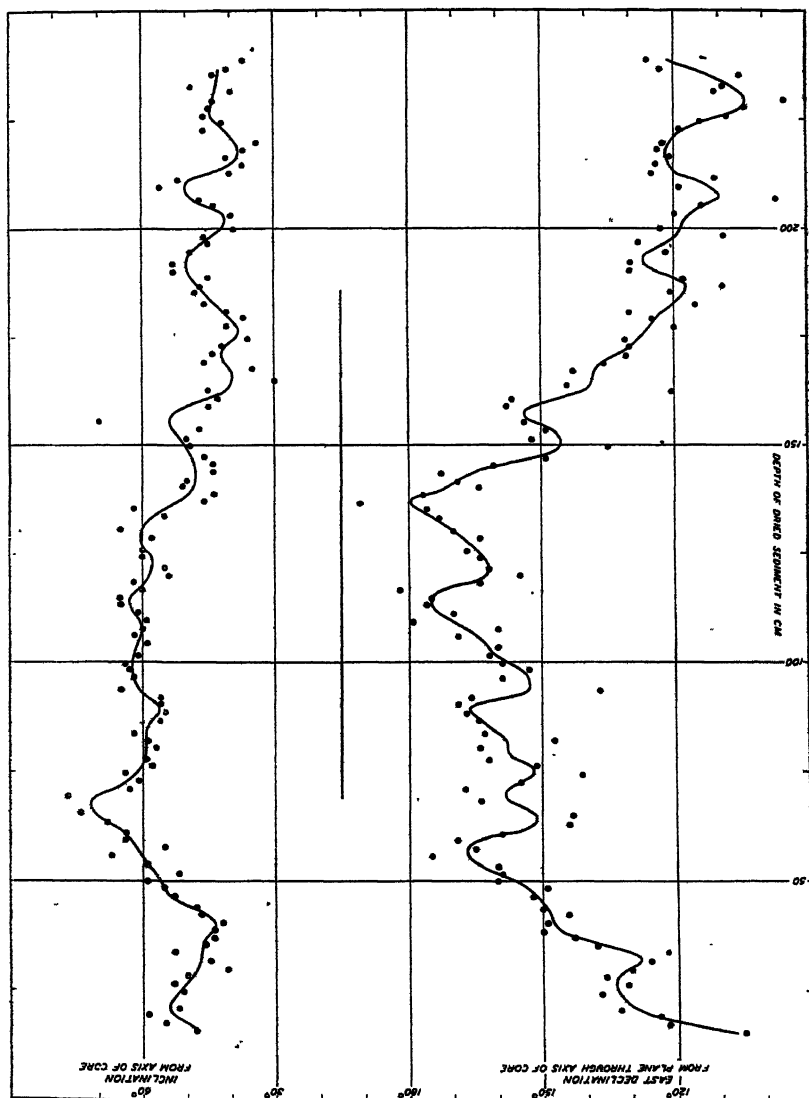
Figure 11. Direction of magnetization of sediments in core No. 3 (curve smoothed by formula $(a+2b+3c+2d+e)/9$).

artificial magnetic fields. This equipment (figure 10) permits detection of a magnetic moment as minute as that of a good steel magnet one millimetre long and one-hundredth of a millimetre in diameter. Further refinements will increase many-fold the sensitivity of this method.

Tests on one core show distinct differences in the direction of magnetization at various depths (figure 11). If it is found that these differences are the result of

differences in the Earth's field at the time the sediments were deposited, we have in the core-samples a means of extending the past history of geomagnetism and its secular changes.   Interrelating individual cores by similarities in their magnetism would determine more accurately than surface observations the extent and character of the regional magnetic anomalies of the ocean bottom.

With improved knowledge of magnetic anomalies, the question of interrelations with gravimetric anomalies and deep-focus earthquakes may be answered.   It is frequently noted that gravimetric and magnetic anomalies occur in the same regions.   Apparently areas of deep-focus earthquakes are confined to the great regions of positive anomalies of magnetic vertical intensity around the Pacific Ocean.   Vening Meinesz has found an excess of gravity in the oceans and a sudden decrease at the continental borders, and has demonstrated that these deviations of isostasy cannot be situated in the Earth's thin crust but must extend far into the plastic layers below.   There must result mass flow in the Earth's interior, downward below the oceans and upward below the continents.   The charts of isoporic foci and their motions with time certainly lend strength to the idea, indicating the interior of the Earth to be more mobile than the external layers, not only as a whole but regionally.   May not, then, the anomalies of gravity and of geomagnetism and the deep earthquakes have a common origin in mass flows in the inner Earth ?   Thus the facts of the geophysical observer of geomagnetism in the field and in the laboratory may contribute much to a complete picture of the Earth's crust and particularly of that great portion which is the ocean bottom.

§ 5. RECENT DERIVATIONS OF SECULAR CHANGE

The Department of Terrestrial Magnetism has compiled isoporic charts of the geomagnetic field for the four epochs of 1912·5, 1922·5, 1932·5, and 1942·5 for declination, inclination, and the northward, eastward and total components of intensity.   These were compiled preparatory to the construction of new world isomagnetic charts of these elements in seventeen sections for the epoch 1945·0.

All available repeat-station observations have been used in compiling secular change for each of the seven elements and components.   All have been corrected for the influence of extraneous geomagnetic variations, including those related to the sunspot cycle (not properly considered as secular change), post-perturbation, annual variation, and the quiet-day solar daily variation.   Graphs of the observed values so reduced to " normal " were made for each station.   Specimen graphs for the station Suez and the Alibag Observatory are reproduced in figures 12 (*a, b, c*) and 13 (*a* and *b*).   The values as actually observed are plotted in order to show the order of correction required to obtain reduced " normal " values.   The corrections at Suez are largest for the horizontal and northward components, due to the considerable latitude-effect of the geomagnetic variation with sunspot-cycle and of post-perturbation.   The corrections are of greatest percentage benefit when the secular changes are small.   The same considerations apply also in the case of observatories such as Alibag; here the corrections (figures 14 *a* and 14 *b*), as for many series of data, have been of material assistence in improving estimates of secular change in intensity near 1910.

Graphs of the foregoing kind were scaled at epochs 1912·5, 1922·5, 1932·5, and 1942·5, and plotted on world maps.   A sample of these is given in figures 15 (*a, b, c*) and 16 (*a, b, c*), showing the isopors in declination and horizontal

Figure 12 a. Specimen of estimated geomagnetic departures, mean of hour (MH) from mean of year (MY₀) centred about mean of month (MM₀), including yearly departure owing to variation with sunspot-cycle (RV) for Suez, station C of 1935.

COMPUTATION OF GEOMAGNETIC DEPARTURES,   MEAN OF HOUR (MH) FROM MEAN OF YEAR (MY₀)   CENTERED ABOUT MEAN OF MONTH (MM₀)

STATION _____ COUNTRY _____ DATE _____

LEGEND

$\delta Y'_{4p}$, $\delta Y'_{qp}$, and $\delta Z'_{qp}$ = Tabular mean hourly values geomagnetic north, east, and vertical components of S_q averaged around parallels geographic latitude

$\delta Y'_q$, $\delta Y'_q$, and $\delta Z'_q$ = Geomagnetic north, east, and vertical components of S_q

f_q = Tabular daily multiplicative factor yielding estimated daily value of S_q from its monthly mean value

$\delta Y'_{q_h}$ and $\delta Y'_{Q_h}$ = Mean hourly values geomagnetic north and east components of departures of S_q from corresponding values of S_q averaged around parallels geographic latitude

W, E, and S = Corresponding tabular mean hourly departures $\delta Y'_{q_h}$ or $\delta Y'_q$, averaged for winter, equinox, and summer for Greenwich hour of observation

w, e, and s = Tabular monthly multiplicative factors for W, E, and S, yielding products whose mean sum is estimated corresponding mean monthly departure of $\delta Y'_q$ or $\delta Y'_q$

AV_{Xy} and AV_{81} = Symmetrical and sinusoidal parts of annual variation

f_{xy} and f_{81} = Corresponding multiplicative factors yielding AV_{Xy} and AV_{81}, in any geomagnetic latitude

f_p = Tabular multiplicative factor yielding P in any geomagnetic latitude

$\delta Y'_p$ and $\delta Z'_p$ = Geomagnetic north and vertical components of P

$\delta Y'_{XY}$, $\delta Y'_{XY}$, and $\delta Z'_{AV}$ = Geomagnetic north, east, and vertical components of AV

P = Daily tabular values geomagnetic post-perturbation as expressed by daily mean departures from monthly mean

Figure 12 b.   Specimen of computation of geomagnetic departures for mean of hour (MH) from mean of year (MY₀) centred about mean of month (MM₀).

MAGNETIC ELEMENTS AND THEIR SECULAR VARIATIONS BY EPOCHS

Figure 12 c. Summary magnetic elements and their secular variations by epochs from 1910, Suez and Helwan.

VALUES OF D, H, AND I

Station ___SUEZ_____   Latitude ___+29°, 56.4'___

Country ___EGYPT, AFRICA_____   Longitude ___32° 29 E___
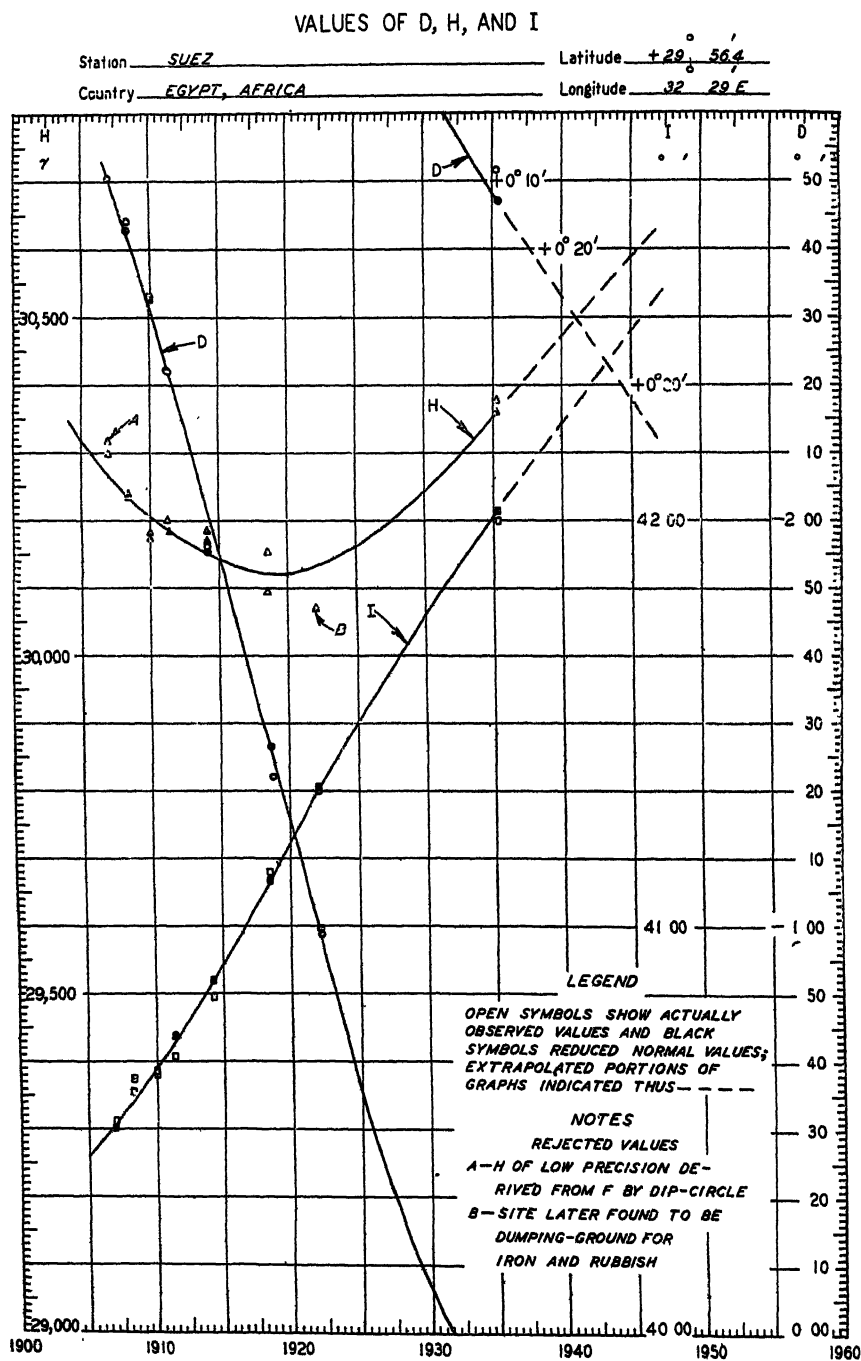
Figure 13 *a*.   Specimen graphs for values of *D*, *H* and *I* from 1905, Suez.
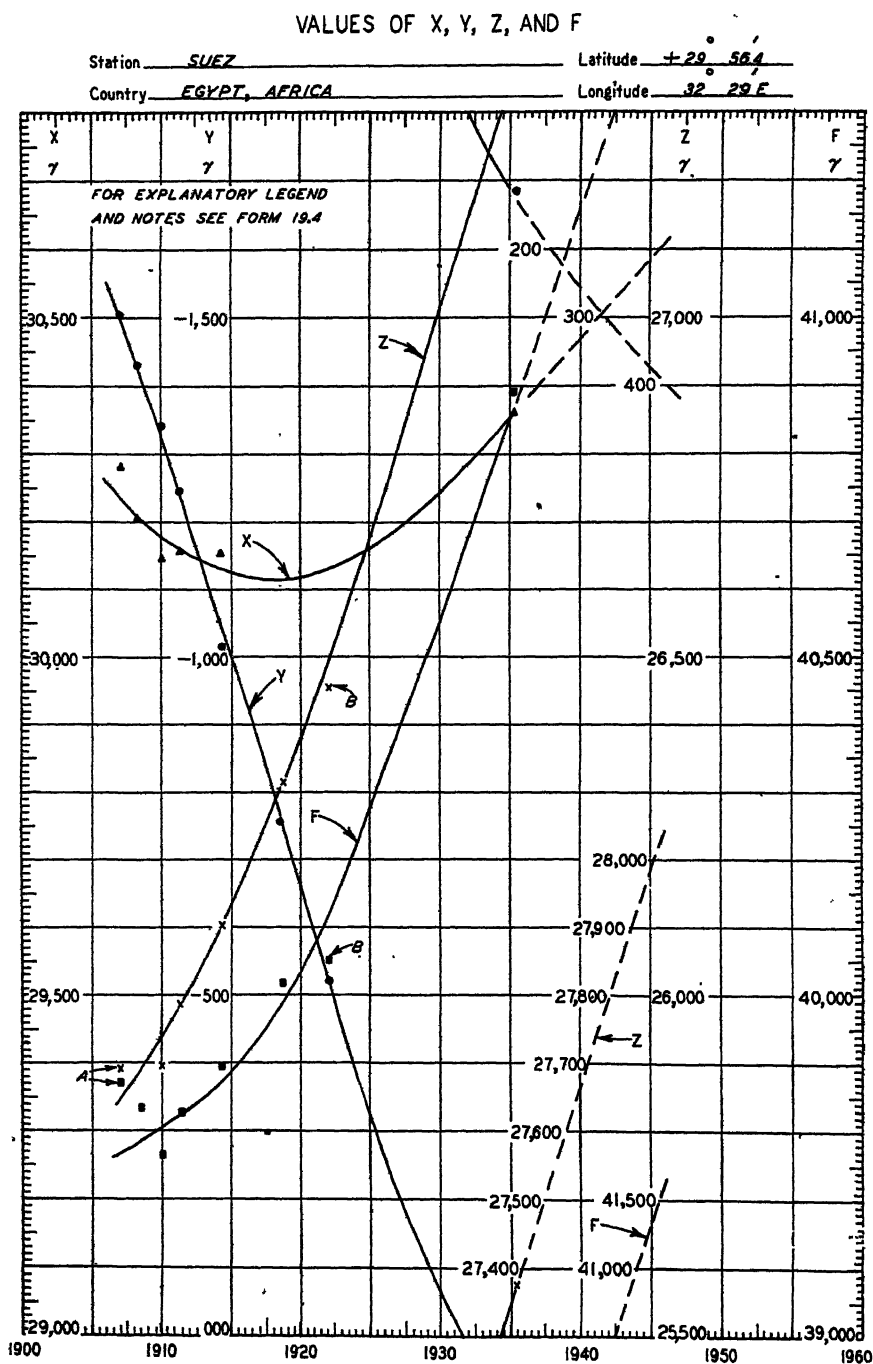
Figure 13 *b*. Specimen graphs for values of *X, Y, Z* and *F* from 1905, Suez.

VALUES OF D, H, AND I

Station ___ *ALIBAG* _____ Latitude __+ *18* ° *38* ′

Country ___ *INDIA, ASIA* _____ Longitude __ *72   52 E* ′

Figure 14 *a*.   Specimen graphs for values of *D*, *H* and *I* from 1905 for Alibag.
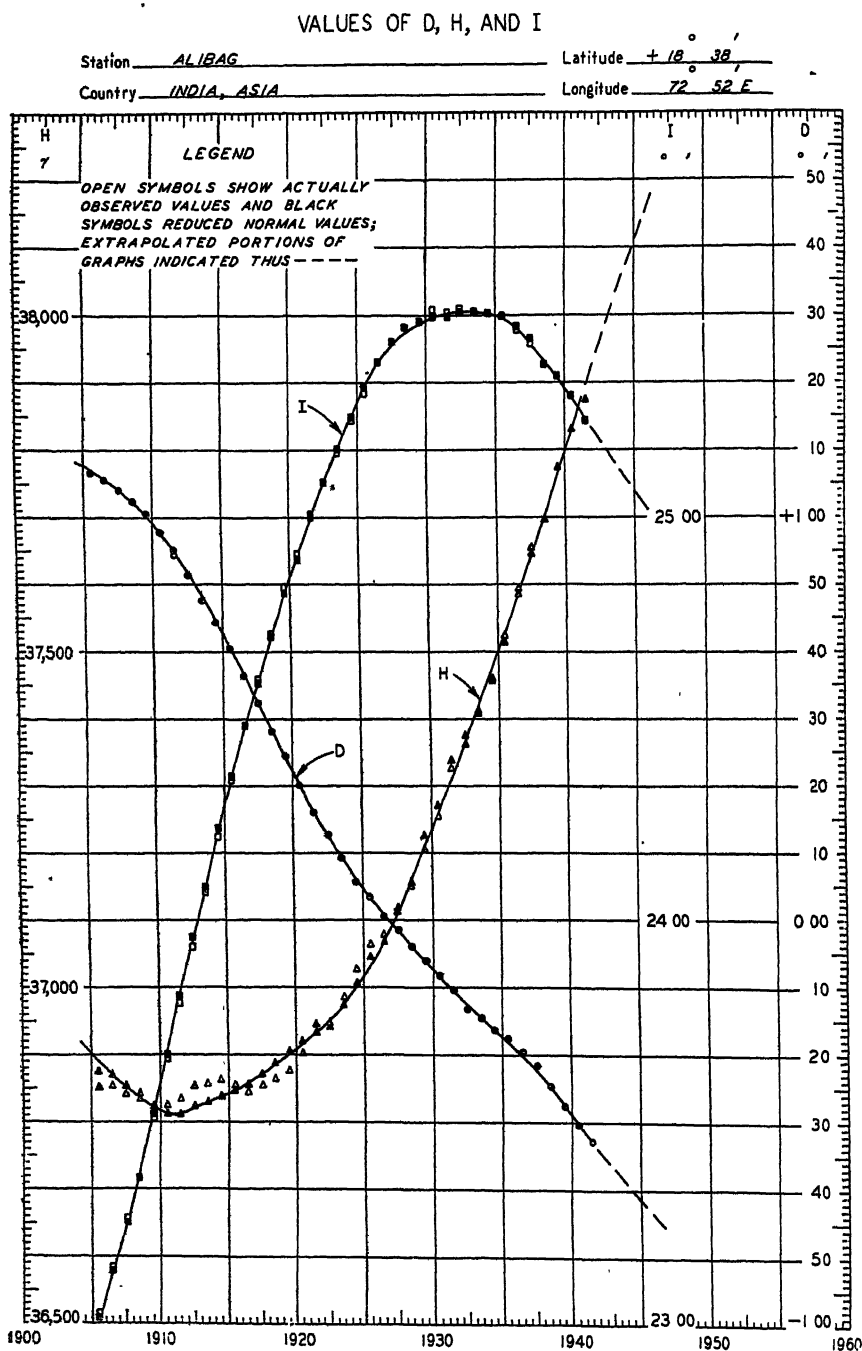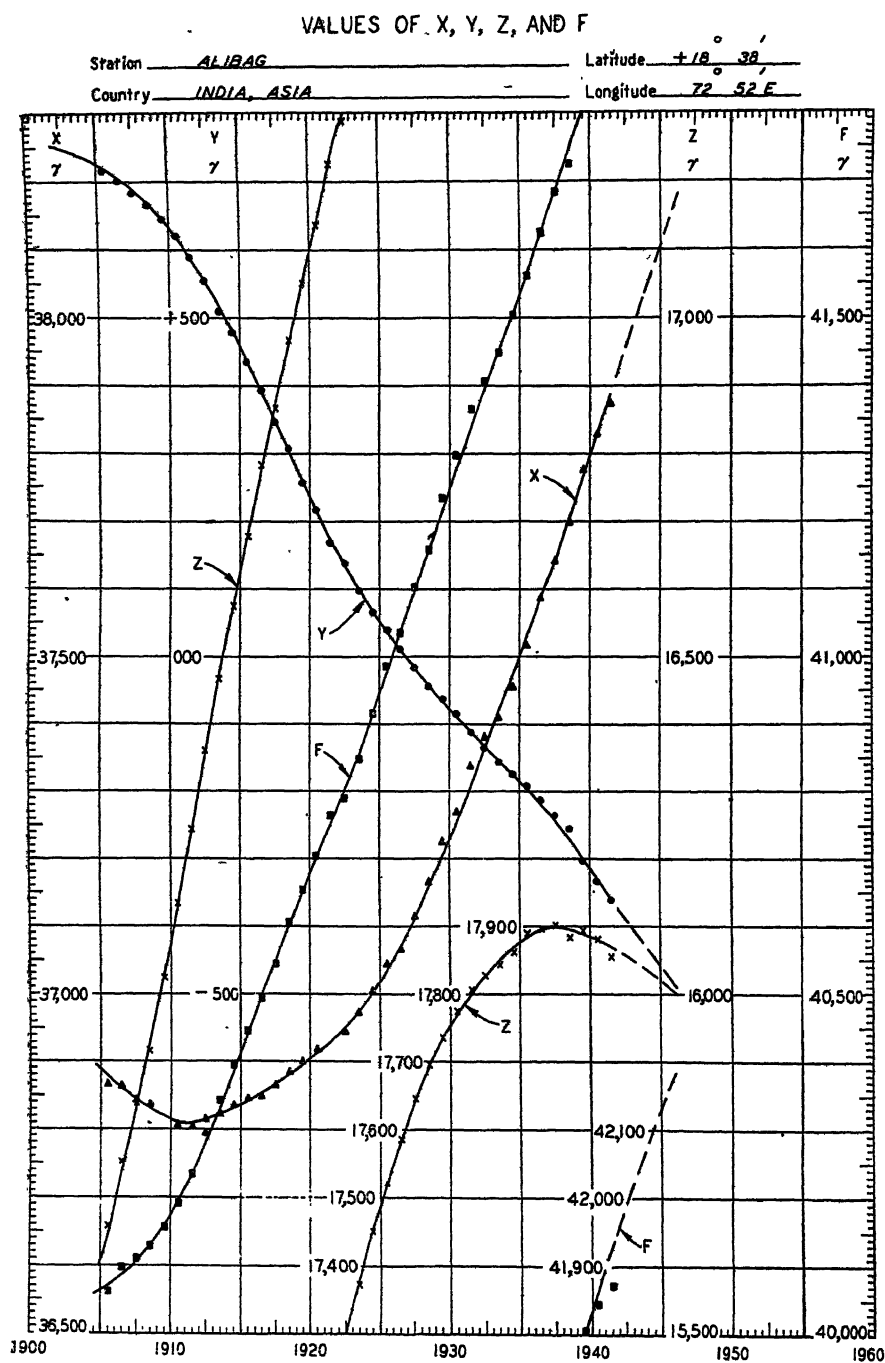
Figure 14 *b*. Specimen graphs for values of *X*, *Y*, *Z* and *F* from 1905 for Alibag.
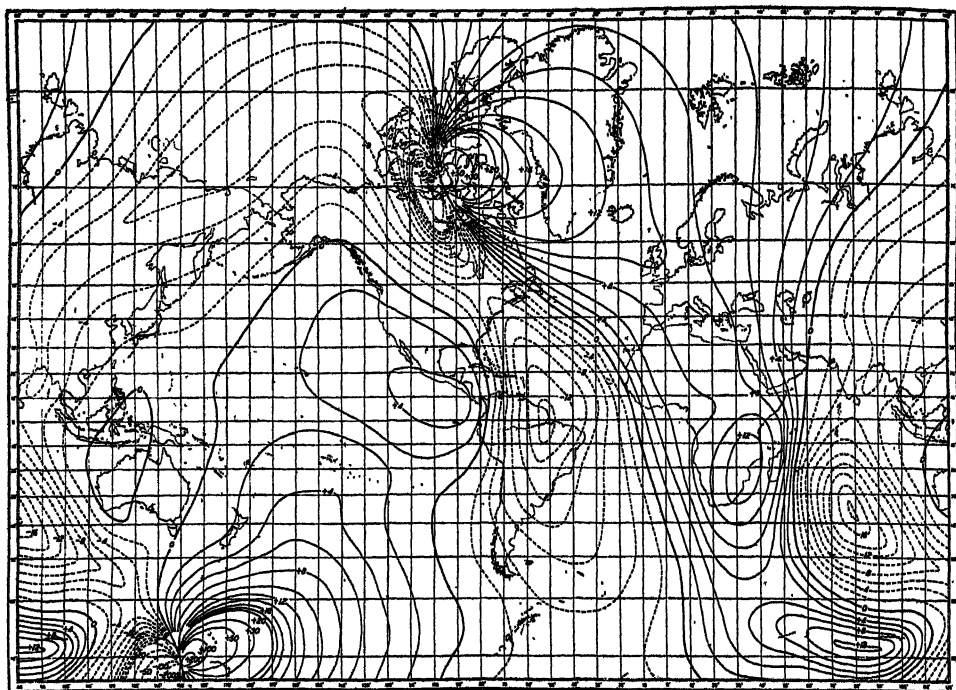
Figure 15 a.   Isoporic chart for declination (lines of equal annual change in minutes of arc per year), epoch 1912·5.
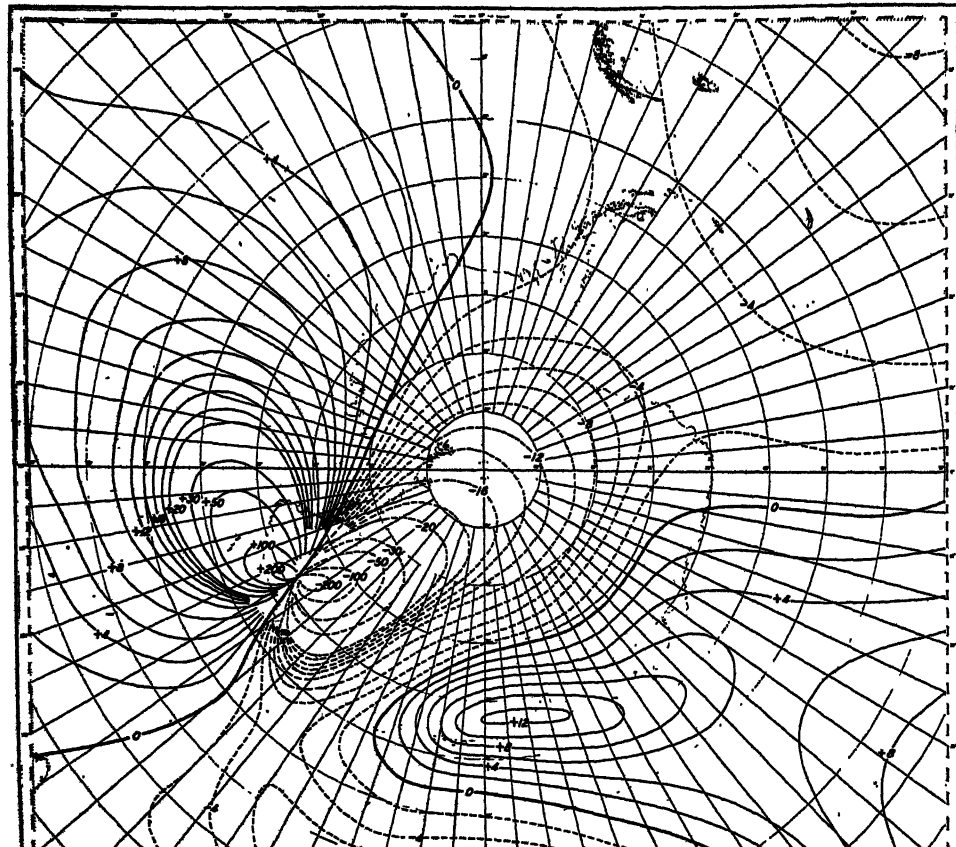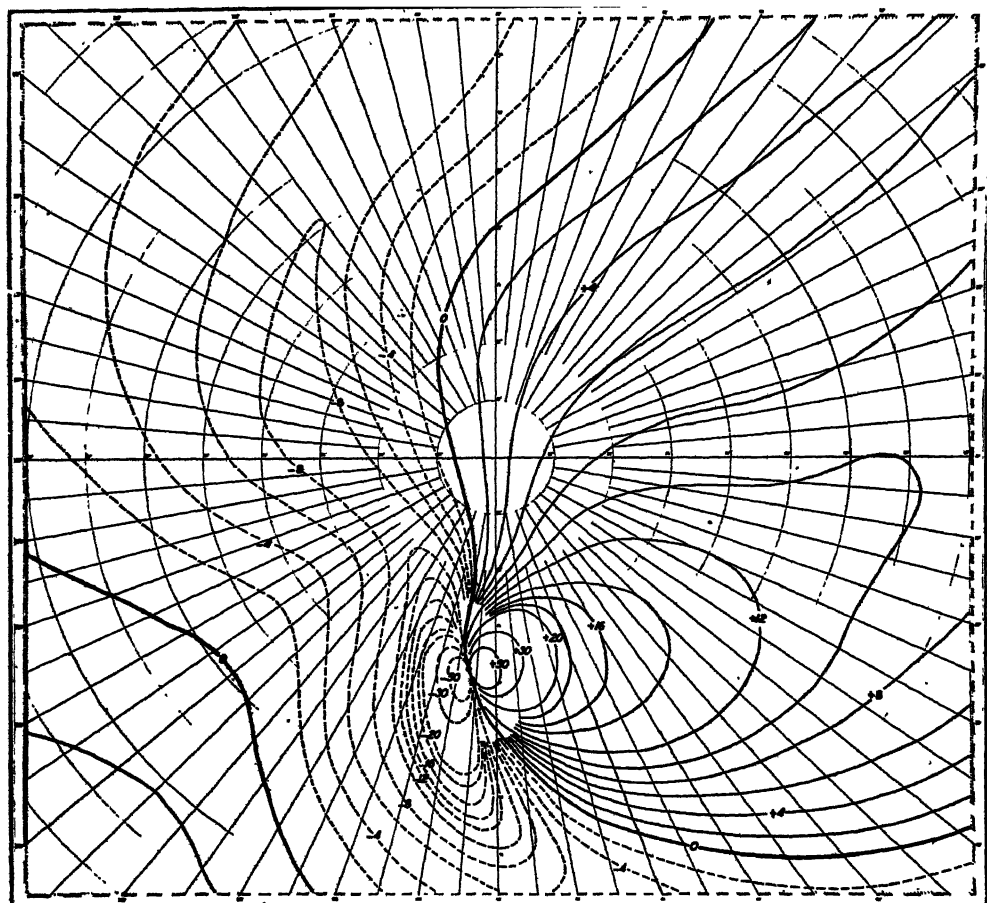
Figure 15 *c*.   Isoporic chart·for declination (lines of equal annual change in minutes of arc per year), South Polar region, epoch 1912·5.
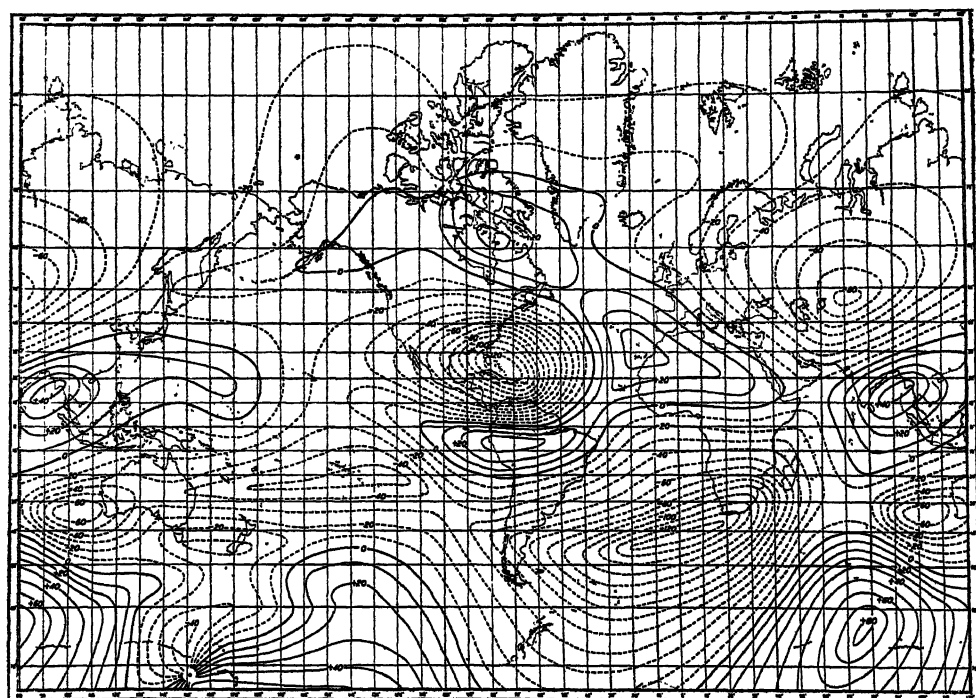
intensity for epoch 1912·5.   These were first drawn independently of one another and then adjusted to mutual consistency such that the vertical component of the curl of the field was nearly zero at the corner of each of the 10°-tessera.   The adjustments were further continued to ensure vanishing of line-integrals of the northward component along meridians and of the eastward component along parallels of latitude.

The isopors near the magnetic poles are based on a theoretical fit of available data, based on a power-series expansion for the potential due to Chapman (1942). The theoretical fit of secular changes near the dip-poles was also used to estimate
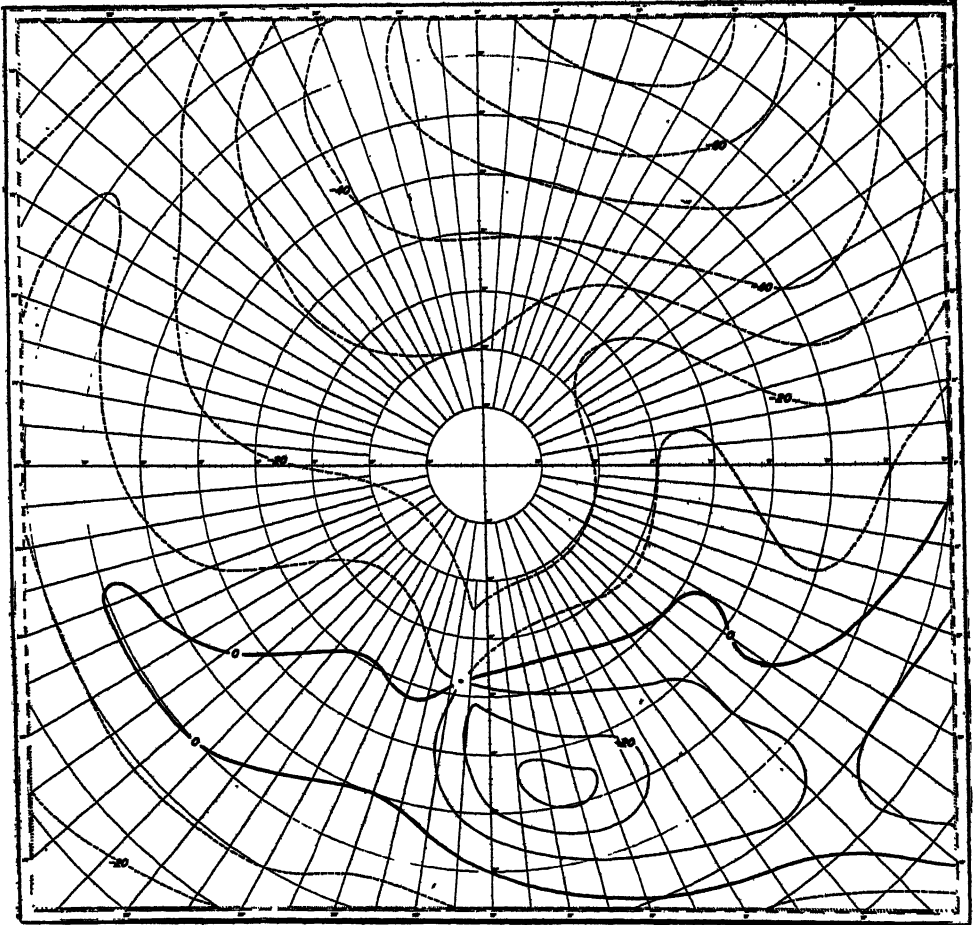
Figure 16 *b*.   Isoporic chart for horizontal intensity (lines of equal annual change in gammas per year), North Polar region, epoch 1912·5.

the positions of the north and south magnetic poles given in table 1.   These tentative estimates assume the revised location of the north magnetic pole by Wasserfall (1939) based on Amundsen's observations of 1904·5 (Graarud and Russeltvedt, 1925) and of the recent flight of the *Aries*.   The spacing between successive positions of the pole at the various intervals of time are adjusted in proportion to the velocities of the pole estimated from secular change.   The position of the south magnetic pole is taken as that recently supplied by Farr (1944), the secular change in horizontal intensity and declination, together with the elliptical configuration of the isoclinics, as found by a survey under Mawson in 1912 (Webb, 1925).
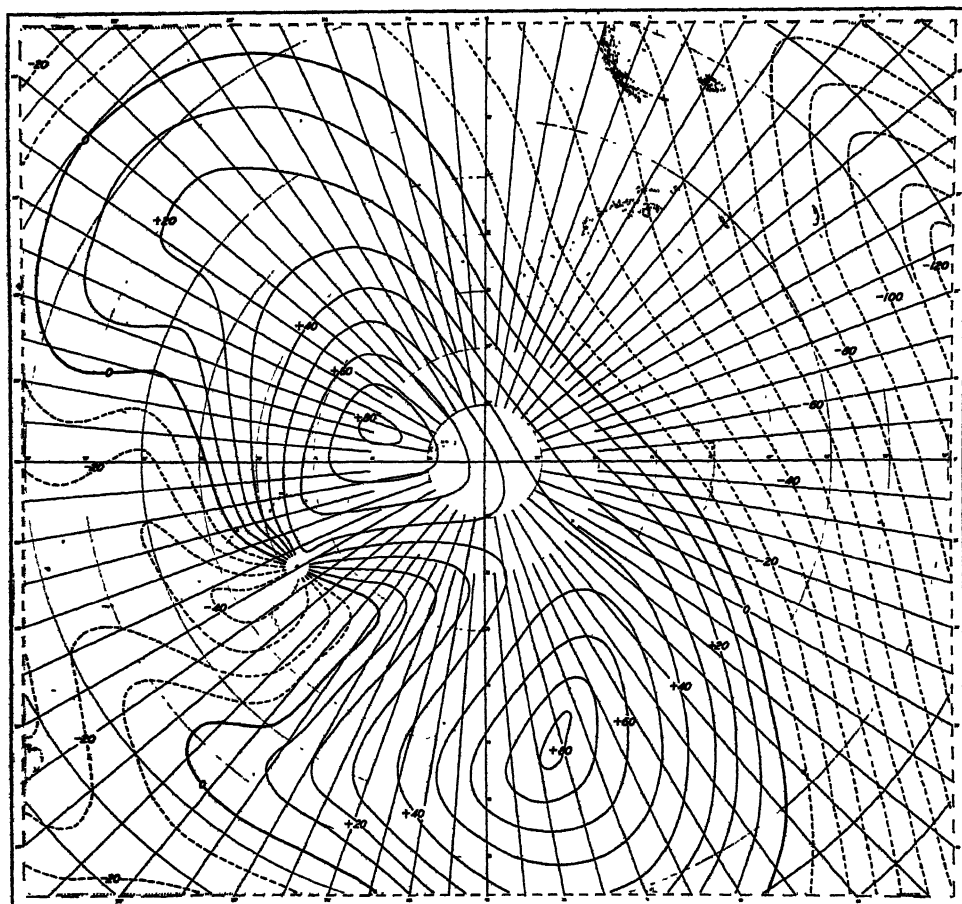
Figure 16 *c*. Isoporic chart for horizontal intensity (lines of equal annual change in gammas per year), South Polar region, epoch 1912·5.

Table 1. Coordinates of north and south magnetic poles.

Date or epoch	Based on secular-change charts for four epochs and positions observed in 1904·5 for north pole and 1912·5 for south pole				Based on isomagnetic charts for four epochs and positions observed in 1904·5 and 1945	
	North latitude	East longitude	South latitude	East longitude	North latitude	East longitude
	°	°	°	°	°	°
1904·5	70·5*	263·5*	....	....	70·5*	263·5*
1912·5	70·9	263·2	71·2†	150·5†	71·3	262·7
1922·5	71·4	262·8	70·2	149·2	72·6	261·8
1932·5	71·9	262·4	69·0	148·1	74·0	260·0
1942·5	72·6	262·1	68·3	146·2	75·4	258·2
1945·0	72·8	262·0	68·2	145·4	76·0‡	258·0‡

* After Wasserfall (1939), based on observations by Amundsen.
† After Farr (1944), based on observations by Mawson.
‡ Preliminary values communicated verbally as observed on *Aries*.

The isoporic charts show agreement with Chapman's rule that the number of nodes equals the number of foci minus 2 (Chapman, 1942). For declination there are eight nodes and ten foci with no ray-poles. For horizontal intensity there are ten nodes, ten foci, and two ray-poles, the latter counting as foci in the application of the rule. New features also depicted for the first time are the distribution of lines near the magnetic poles.

The ten-year successive positions of positive and negative foci in declination and horizontal intensity for the four epochs ten years apart are plotted in figure 17. The most notable feature is the large and systematic departures in position per ten-year period. Tables 2 and 3 list the locations of foci, their magnitudes, and
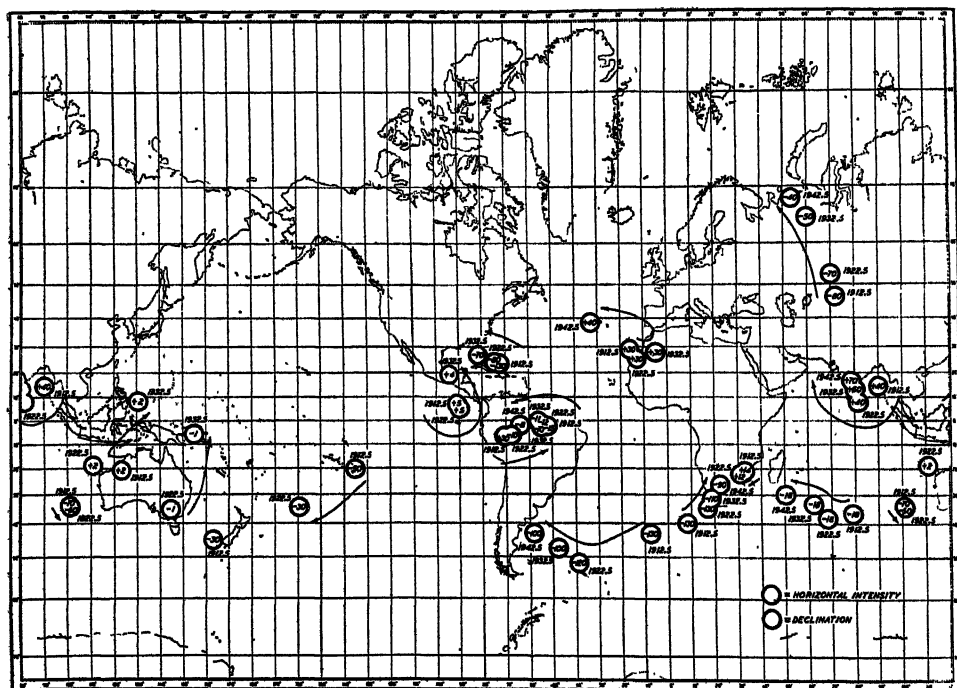
Figure 17. Successive approximate ten-yearly positions and magnitudes of isoporic foci; declination in minutes of arc and horizontal intensity in gammas ; epochs 1912·5, 1922·5; 1932·5 and 1942·5.

positions. In the Eastern Hemisphere the foci of the secular change in horizontal intensity show in general circulation in position to the north and to the west. This finding is in agreement with the general results for the position of the magnetic equator indicated in figure 3. In the Western Hemisphere there seems some indication of motion to the south. The foci for secular change in declination, on the other hand, seem to show motion in earlier years which is in certain cases towards the east, there being less regularity than that noted for the secular change in horizontal intensity.

These new isoporic charts offer several new and interesting possibilities. Although constructed independently of one another, the differences of the chart, epoch by epoch, reveal surprisingly simple consistent patterns. These new patterns represent the time-rate of change of secular variation per ten-year

1912.5			1922.5			1932.5			1942.5		
Latitude	Longitude, E	f	Latitude	Longitude, E	f	Latitude	Longitude, E	f	Latitude	Longitude, E	f
°	°	′	°	°	′	°	°	′	°	°	′
−21	33	+14	−23	31	+12	...	...	−18	...	...	...
−36	79	−18	−38	68	−18	−33	62	+2	−30	50	−18
−21	123	+2	−19	111	+2	+7	130	+1	...	...	...
...	...	...	−35	144	−1	−5	155	+4	...	...	...
+7	269	+5	+5	270	+5	+19	265	−11	...	...	...
−2	308	−15	−2	306	−12	0	304	...	−2	296	−8

Table 3. Positions and values of maximum and minimum isoporic foci (f) in horizontal intensity for four epochs

Epoch

1912.5			1922.5			1932.5			1942.5		
Latitude	Longitude, E	f	Latitude	Longitude, E	f	Latitude	Longitude, E	f	Latitude	Longitude, E	f
°	°	γ	°	°	γ	°	°	γ	°	°	γ
+47	72	− 80	+53	70	− 70	+65	60	− 50	60	53	− 40
+15	90	+ 40	+ 8	81	+ 40	+13	80	+ 60	+17	79	+ 70
−34	101	− 70	−35	101	− 50	...	...	...	...	...	...
−20	225	− 50	−35	200	− 30	...	...	...	...	...	...
+24	287	−130	+24	285	− 80	+27	278	− 70	...	...	...
−43	352	−130	−52	321	−120	−47	112	−100	−43	301	−100
−40	8	−130	−35	17	−130	−31	19	−110	−26	22	− 90
+ 7	289	+ 30	− 6	292	− 10*	− 3	305	− 10*	...	...	...
+29	344	+ 30	+26	347	+ 30	+28	355	+ 30	+38	328	+ 40
−45	163	− 30	...	...	...	...	...	...	...	...	...

interval. It should prove interesting to try and relate these new values to other geophysical material. The consistency of the differences found are encouraging in another way. They may indicate that the isoporic charts of the northward and eastward components may be successfully subjected, for the first time, to spherical harmonic analysis. It is planned to undertake such a trial analysis, a check being provided by comparing the separate results for the two components. If the analysis is satisfactory, necessary and desirable improvements in the isoporic charts for vertical component can be made, for instance, in polar regions and over the oceans. The extrapolation of the secular-change field or its equivalent in terms of a thin current-distribution at various depths within the Earth is also contemplated.

An interesting hint of the results of this extrapolation can perhaps be derived by noting the considerable complexity of the isoporic charts in horizontal intensity and declination (no attention should be paid to complexities near the magnetic poles). For this purpose we should preferably have the vector combination of the values of secular change in declination and horizontal intensity. The complexity in pattern must increase rapidly downwards in the Earth, and it is highly probable therefore that this extrapolation would attain unreasonable complexity at modest depths, only a moderately small fraction of the distance to the Earth's centre. We thus restrict the probable site of the origin of secular change to smaller spacial dimensions than hitherto, perhaps within the mantle of the crust of the Earth.

§ 6. FUTURE GEOMAGNETIC SURVEYS

The present slow and costly methods of geomagnetic surveys seem likely to be supplemented by new techniques and methods, the possibilities of which are now becoming gradually apparent. Clearly, the old procedure of measuring the Earth's field at different times in different places in one region during one year or years and in another at a different epoch is inadequate to all the needs. What is required is a description of the geomagnetic field during a given year based on measurements made during that year at an orderly set of points sufficiently close together and spaced in a manner nearly independent of topography, areas of land and sea, and climate. The prophecy of Chapman in the first Charles Chree Address (1941) that such measurements might well be made from planes promises early realization. A project of this kind is now feasible instrumentally and affords attractive post-war possibilities in application.

The available data are already of necessity squeezed rather hard to obtain useful isomagnetic charts for the epoch 1945 in some regions.

As stated above, charts of first accuracy could not be prepared without extensive and uncertain extrapolations for secular variations; it was impossible to meet military specifications for accuracy on charts on which our magnetic-mine warfare and degaussing were to be based. Mathematical adjustments were resorted to in making up for the paucity of observations. Believing that a non-potential field does not exist, maps were mutually adjusted so that the curl of the field vanished. It is probable that these adjustments give a more accurate representation of the field than could otherwise be obtained, but no amount of mathematical adjustments can ever compensate for absence of observations.

We have sought to improve our knowledge by increasing the accuracy of field

observations, and this step has been important in accurately determining secular changes. However accurately the magnetic elements are determined at a given point, absence of observations at intervening points leaves our information still inadequate for surface isomagnetic maps unless we can establish almost an infinite number of stations.

This condition may now be remedied by magnetic observations with instruments airborne above the surface of the Earth, and thus farther away from magnetic anomalies which depreciate the accuracy of surface observations. Travelling at speeds which a few decades ago seemed fantastic, continuous observations may now be obtained over a large area in a single day, and in a relatively short time a survey of the normal geomagnetic field can be made. The accuracy of such techniques, thanks to the achievements during the past war, can be kept more than adequate for both practical and theoretical purposes. Electronic circuits, monitored by standard cells, can maintain standard electric currents in flight, so that magnetic profiles, accurate to a few tenths of gammas or better, in absolute value in any component can be obtained. The following limitations in the present airborne gear must be overcome:—(a) Stable operation must be provided over extended periods of time of measurements of intensity: this is synonymous with negligible base-line and scale drifts in the measuring circuits. (b) Modifications are necessary to permit measurement of the angle of declination, that is, the angle between the horizontal projection of the detector coil at any instant and true north or the true heading of the aircraft. (c) Observations of the angle of dip will involve the construction of a suitable gyro mechanism to establish an accurate horizontal plane; among problems in this connection are the elimination of centrifugal pseudo-gravity effects due to plane manoeuvres or bumpy air.

The advantages of an effective airborne surveying instrument are great, and include:—(a) Ability to cover large areas quickly and continuously. (b) Elimination of purely surface anomalies, which are of no significance in the preparation of charts of the normal field; thus, by flying at altitudes of, say, 20 000 feet, smoothed values of the geomagnetic elements are obtained directly, without the necessity of mathematical treatment or judgment on the part of the person processing the data of many agencies and of various degrees of accuracy. (c) Ability to define uniquely the magnetic field by flights in two or more horizontal planes.

Perfection of a suitable airborne instrument for magnetic surveying will require both considerable time and funds to overcome these limitations. Nevertheless, the end results possible would justify such expenditure of both time and funds.

Azimuths for declination can be obtained by automatic or manual observations of the Sun, smoothed by gyroscopic stabilization. An example of what can be done with airborne methods is supplied by the recent flights (August 1945) of the *Aries* under sponsorship of the Imperial School of Navigation on the initiative of the Astronomer Royal. In two flights of less than 24 hours over entirely different regions, a new position of the magnetic pole was determined, probably more accurately than ever before. This achievement may be compared with the tedious and dangerous expeditions of Ross and Amundsen to accomplish the same purpose many years ago. The instruments used were magnetic compasses of various types, including gyrostabilized compasses, and also a three-element, saturated-core type of magnetometer. The last, of a relatively low order of accuracy, was not designed for the flights; it had to be held level manually during

an observation. Furthermore, only one element could be observed at a time, and there was no means for continuously recording—thus truly average values of the observations obtained seem consistent. This pioneer work demonstrates what may be expected when magneticians take full advantage of the technical progress of the last five years.

§7. CONCLUSION

During the war, as by-products of military demands, many improved techniques, applicable not only for surveys and experiments but also for mass reductions, analyses, and correlations by machine methods, have been developed. Peace-time applications of these must serve both continuation of surface and initiation of aerial magnetic surveys by our several Governments, as well as interpretative investigations of the vast existing stock of data and that potentially possible in much greater degree in the near future. It is to be hoped that a long period of amity and cordial relations between all nations is now on the horizon. May it give firm foundation in forwarding and coordinating large activities in Earth physics based on that truly scientific principle of entire lack of desire for personal or national aggrandizement.

Finally, such progress requires the combined efforts of the physicist, the mathematician, the geologist, the astrophysicist, and the engineer to enable the geophysicist to solve the riddles of the Earth's main magnetic field and of its secular variations.

REFERENCES

BERKNER, L. V. and McNISH, A. G., 1938. "The ephemeral variations of the Earth's magnetism." *Carnegie Inst. Wash.*, Pub. No. 501, 223.

CHAPMAN, S., 1941. "Charles Chree and his work on geomagnetism." *Proc. Phys. Soc.*, **53**, 629.

CHAPMAN, S., 1942. "Notes on isomagnetic charts.—VII. Mathematical notes on isoporic charts and their singular points." *Terr. Mag.*, **47**, 115.

CHAPMAN, S. and BARTELS, J., 1940. *Geomagnetism* (Oxford : Clarendon Press).

FARR, C. C., 1944. *Terrestrial magnetism* (Adelaide : The Hassell Press). (B.A.N.Z. Antarctic Research Expedition, 1929–1931, under the command of Sir Douglas Mawson. *Reports*, Series A, 4 (Meteorology and terrestrial magnetism), Part I).

FISK, H. W., 1929. "Secular variation of magnetic intensity and its accelerations in Pacific countries." *Proc. Fourth Pacific Sci. Cong., Java*, 517.

FISK, H. W., 1932. "The unsymmetrical distribution of the magnetic secular variation." *Terr. Mag.*, **37**, 235.

FISK, H. W. and FLEMING, J. A, 1931. "On the distribution of permanent repeat-stations." C.-R. Assemblée de Stockholm, Août 1930 ; *Union Géod. Géophys. Internat., Sec. Mag. Electr. Terr.*, *Bull.* No. 8, 293.

FLEMING, J. A., 1937. "Magnetic surveys of the oceans." Washington : *International Aspects of Oceanography*, p. 50.

FLEMING, J. A., 1938 a. "The general magnetic field of the Earth and its secular variation." *Carnegie Inst. Wash.*, Pub. No. 501, 205.

FLEMING, J. A., 1938 b. "Terrestrial magnetism and oceanic structure." *Proc. Amer. Phil. Soc.*, **79**, No. 1, 109.

FLEMING, J. A. (Editor), 1939. *Terrestrial magnetism and electricity* (New York : McGraw-Hill Book Co.). See also FLEMING, J. A., 1940. "Trends of research in terrestrial magnetism and electricity." *Trans. Washington Meeting*, Sept. 1939 ; *Internat. Union Geod. Geophys., Ass. Terr. Mag. Electr.*, *Bull.* No. 11, 41.

GELLETICH, H., 1937. "Ueber magnetitführende eruptive Gänge und Gangsysteme im mittleren Teil des südlichen Transvaals." *Beitr. angew. Geophys.*, **6**, 337.

GRAARUD, A. and RUSSELTVEDT, N., 1925. "Die erdmagnetischen Beobachtungen der Gjöa-Expedition 1903–1906." *Geofys. Pub. Oslo*, **3**, No. 8.

PIGGOT, C. S., 1938. "Core samples of the ocean bottom and their significance." *Sci.*

Dr. JOHN A. FLEMING,
Third (1945) Charles Chree Medallist.
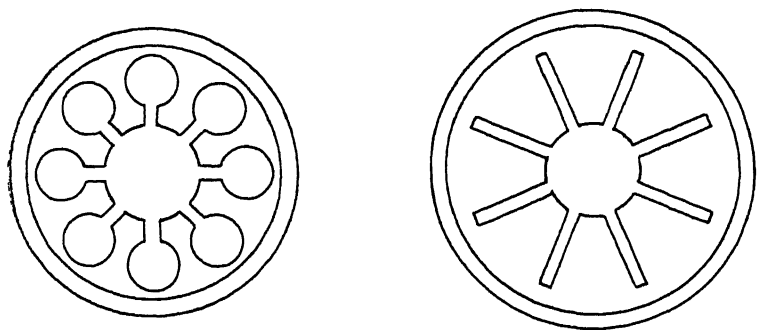
Figure 1. Cavity magnetron resonator systems : sectional view.
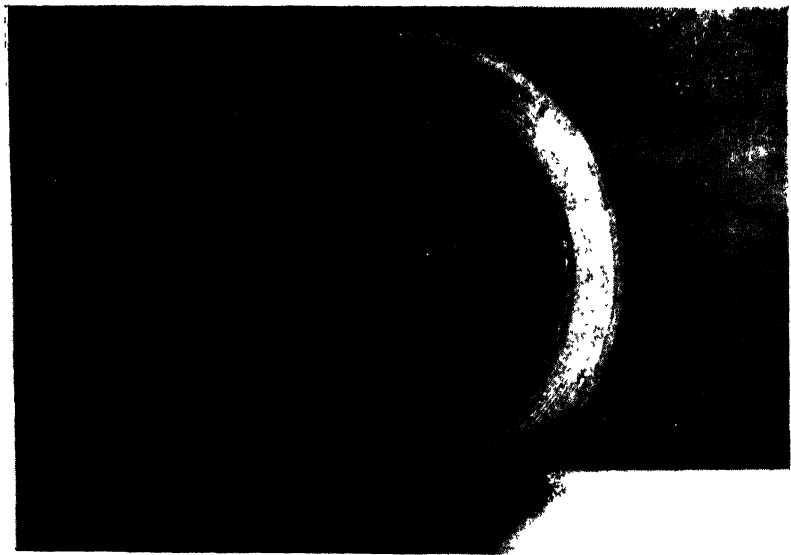
Figure 2. Anode block of one of the earliest cavity magnetrons made in the University of Birmingham.
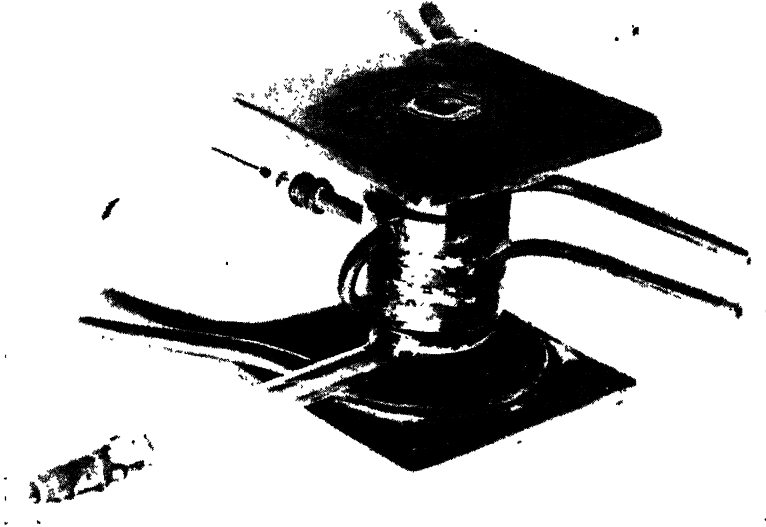
Figure 3.   The original cavity magnetron detached from pumping system.
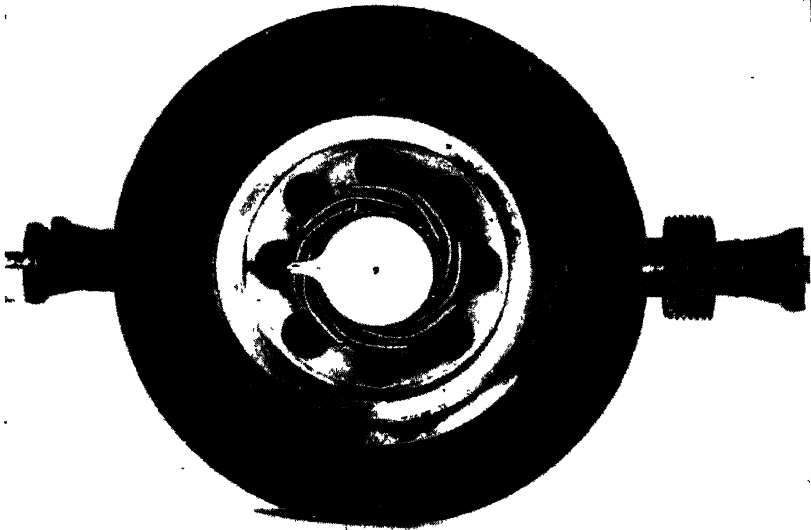
Figure 4.   The CV 64, a 9–10-cm. echelon-strapped magnetron much used by the Services.
Peak power output ∼35 kw.
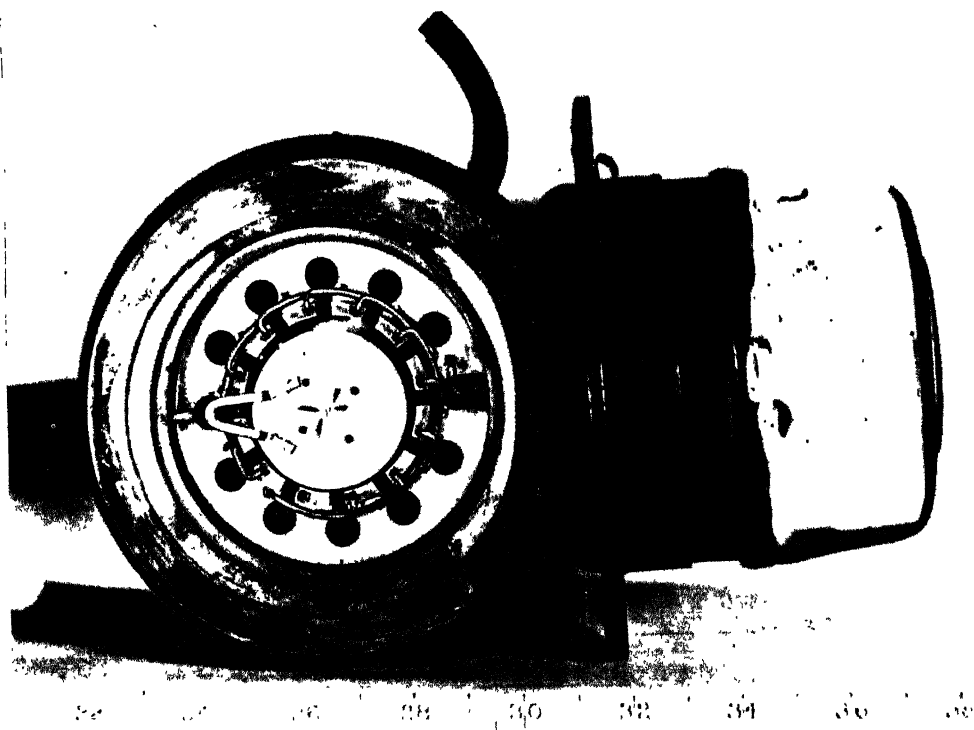
Figure 5.   Megawatt magnetron with end-plate removed to show strap construction.   (Scale in cr
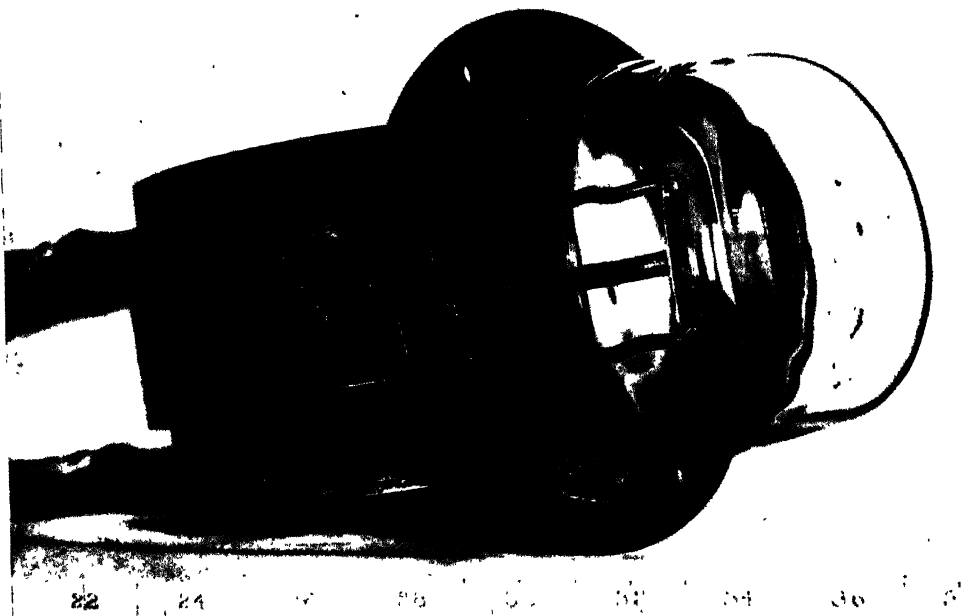
Figure 6.   Megawatt megnetron; second view, showing wave-guide output system.   (Scale in cr

SCHONLAND, B. F. J., 1943. "Thunderstorms and their electrical effects." *Proc. Phys. Soc.*, 55, 445.

SCHRÖDINGER, E., 1943. "The Earth's and the Sun's permanent magnetic fields in the unitary field theory." *Proc. R. Irish Acad.*, A, 49, 135.

SVERDRUP, H. U., 1939. *Physics and geophysics. With special reference to problems in physical oceanography* (Berkeley : Univ. California Press).

WASSERFALL, K. F., 1939. "Studies on the magnetic conditions in the region between Gjöahavn and the magnetic pole during the year 1904." *Terr. Mag.*, 44, 263.

WEBB, E. N., 1925. "Field survey and reductions of magnetograph curves." Australasian Antarctic Expedition, 1911–1914, *Scientific Reports*, Series B, 1 ; *Terrestrial magnetism*, Part I, 1–197. (Sydney).

THE CAVITY MAGNETRON *

BY J. T. RANDALL, F.R.S.,

University of St. Andrews

IN accepting the great honour which you and your Council have bestowed on me I wish to offer my most grateful thanks. I have very little to guide me by way of precedent on this occasion. Either by accident or design the remarks of previous recipients of the Medal have but rarely been reproduced *in extenso*. I must therefore chart my own course. Before I refer to the work you have singled out for special mention I would like to say this:

Not only do I owe a great deal to the laboratories in which I have worked and to their Directors, but I have been specially fortunate in the young people who have collaborated with me. If I had not been trained in Manchester under W. L. Bragg, I might not have developed the interest in crystals which was so valuable as an approach to the problems of luminescence. If I had not spent some years in the Wembley laboratories my approach to valve problems would have been inadequate when I came to think of short-wave generators. And, indeed, if I had not been in the University of Birmingham when Oliphant started his great drive on centimetre-wave work, I should probably have had nothing at all to do with such instruments. I say these things because I believe that at any given time one's past experience affects one's approach to a problem, and determines more than many people think the way in which a problem is solved.

The luminescence work, which started a good deal earlier than that on high-frequency generators, was of two kinds: technical and fundamental. I might mention on the technical side the discovery, by McKeag and myself (1936), of zinc beryllium silicate activated by manganese as a phosphor suitable for excitation by a low-pressure mercury discharge. This phosphor is now a component in many of the mixtures used in the low-voltage fluorescent lamps (see Jenkins, 1945). I think, also, we were among the first to elucidate the technical processes required to obtain a number of high-quality silicate and phosphate phosphors.

On the fundamental side I should like to mention the work of Randall and Wilkins (1946) on the mechanism of afterglow or phosphorescence in different

---

* This short title has been chosen since it covers the greater part of the subject-matter of the Address, although some reference is made to work on the luminescence of solids.

types of material. In a series of recent papers it is shown how the characteristic thermoluminescence is related to electron traps in the solid; further, that the temperature at which light is emitted is a measure of the depth of the trap below the conduction band in the solid. From a knowledge of the relative numbers of electrons in traps of different depths it was possible to calculate the type of decay law obeyed over a narrow temperature range. Wilkins and Garlick (1945) have followed this work with a paper in which many earlier misconceptions concerning bimolecular decay laws have been cleared away. There is one important problem in this field which someone may wish to investigate: the nature of electron traps in particular cases. No direct attack is possible for obvious reasons. The detailed study of phosphors with more than one activator—rare-earth ions would probably be suitable—will probably give information, not only with regard to the luminescence centre, but about the trap, and optical and thermal excitation energies.

The work in which I suspect you are more interested, however, is the discovery by Dr. Boot and myself of the resonator magnetron in the winter of 1939–40. Later, in 1941, Dr. Sayers made an invaluable fundamental contribution. As the magnetron was immediately successful as a short-wave generator it was taken up for the Government by the large electrical firms, and I should like to pay tribute to the work which was carried out by the staffs of the B.T.-H. and the G.E.C. in this connexion.

As the story has not really been told of how this work arose I should like to spend a little time on the origins of the magnetron as we now know it, and to follow this with a brief reference to later experimental developments. At the close of my talk I hope to show a powerful magnetron in operation, and for generous help in this I wish to thank the Admiralty Signal Establishment.

In the early autumn of 1939, with others whom Professor Oliphant had collected together for this purpose, I began to spend some time on problems connected with centimetre waves, as many other physicists were doing elsewhere. The chief problem was to produce a high-power centimetre wave oscillator. The centimetric nature of the wave was required for direction of beam, and the power for long range. At first Boot and myself spent a few weeks studying the Barkhausen-Kurz tube as a detector device; the greater number of workers in the laboratory, however, were concerned with the klystron (Varian and Varian, 1939), both as an oscillator and as an amplifier, and we were naturally interested in the whole field. The klystron was outstandingly important in that it used for the first time closed, or essentially closed, resonators, fitted with grids through which an electron beam passed, for the production of high-frequency power. Such resonators had been considered by Rayleigh in 1897, and the papers of Hansen (1938, 1939) showed how enclosures of this kind lend themselves to certain modes of electrical vibrations. On any given mode the frequency is largely, if not entirely, dependent on the dimensions of the resonator. When such resonators are constructed of copper, three important features are evident:— (i) low h.f. losses; (ii) wave-length stability; (iii) potential capability of large heat dissipation.

To be set against this, however, is the serious difficulty of getting sufficient power into the electron beam, which is necessarily of small cross-sectional area. We began, therefore, to think of other possibilities. Was there any means of using the desirable features of cavity resonators, and at the same time avoiding the limitations of the electron beam ?

Conventional designs of magnetrons were equally notable in failing to meet the desirable features, and by November 1939 we had decided to try to devise a type of magnetron which did so. The Hansen papers already referred to were by this time available in the laboratory, but it was obvious that the types of resonator considered—hollow spheres, cubes, etc.—could not be associated with the cylindrical anode and cathode of the magnetron. The only other types ' of short-wave resonator circuit with which we were familiar were Hertz's original loop-wire resonator and a short-circuited quarter-wave line. A three-dimensional version of the first of these gives a cylinder with a slot down one side, and the second becomes a parallel-sided slot $\lambda/4$ deep (see figure 1). Either design was well suited to the arrangement of a number of segments with cylindrical symmetry round a central cathode, and it is also clear that a combined anode and resonator system could be machined from solid copper to allow of good heat transfer and, it was hoped, high-power dissipation.

No theoretical calculations of the wave-length to be expected from a resonator of the cylinder-slot type had been made, but Hertz had shown, and Macdonald (1902) had confirmed theoretically, that the wave-length to be expected from a simple Hertzian dipole was $\lambda = 7.94d$, where d is the diameter of the ring.

Six resonators, 1.2 cm. in diameter, with slots 0.1×0.1 cm., were employed, and the length of the resonator was 4 cm.; it was hoped that the wave-length of this system symmetrically surrounding an anode-hole of the same diameter would be approximately 10 cm. Preliminary calculations had shown that this design might be expected to operate reasonably well with an applied potential of 16,000 volts and a magnetic field of approximately 1000 oersteds, so that provision of these essentials was a necessary preliminary. In order to save time a D.C. power-supply system was built up, and not a modulated supply, and a large electromagnet was used to give the required field parallel to the cylindrical axis of the valve. Figure 2 shows a view of the first anode block, figure 3 the first resonator magnetron, continuously operated on the pump with glass-metal junctions closed with sealing wax. For various reasons the first trials were delayed, and it was not until 21 February 1940 that the valve was tested. It worked immediately. By various rather crude means it was shown that the output was about 400 w. The wave-length was measured by means of Lecher wires and shown to be 9.8 cm. After this, things moved rapidly, and I will not enumeiate the various technical stages of development, which included the making of sealed-off valves by ourselves and by G.E.C., Wembley, incorporating the famous gold-seal device as a means of fixing the end-plates to the valve. We also received great help on the technical side from S. M. Duke, formerly of the Wembley laboratories. Later, Professor Oliphant organized a small production unit for experimental magnetrons, and during its life over a thousand experimental valves were produced.

There have been many stages in the development of the magnetron since 1940, and it is obviously impossible to go into these in detail. It was, of course, important to see to what extent power could be increased. The Wembley and Birmingham laboratories simultaneously introduced large oxide cathodes, although both laboratories were aware that the French had been experimenting on these lines with the older types of split-anode magnetron. During the summer of 1940 we were able to show that the cavity-resonator principle was of general application; any fears that might have existed concerning the applicability to larger numbers of resonators and shorter wave-lengths proved groundless.

During July 1940 experiments on a 5-cm. valve were carried out, and a small amount of power was also obtained at 1·9 cm. using 30 slot resonators. A little later, September 1940, a scaled-down version of the original 10-cm. valve (6 resonators) was operated at 3 cm.

In the early design work we were aided by some elementary considerations. For electrons passing close to the anode without grazing it the operating voltage turned out to be

$$V = 2 \times 10^7 \times \pi^2 D^2 / N^2 \lambda^2 \phi^2 \text{ volts,} \qquad \ldots \ldots (1)$$

where D is the anode-hole diameter, N the number of resonators, and ϕ the phase difference between adjacent resonator tongues.

The corresponding expression for the magnetic field is

$$H = 7 \cdot 24 \times 10^4 / N\lambda \text{ oersteds.} \qquad \ldots \ldots (2)$$

These expressions were undoubtedly based on quite inadequate theory; but as they were found extremely useful in the design of new valves they deserve a small place in the record of developments. It was later shown experimentally at Birmingham that the resonator magnetron will operate with higher efficiency at higher magnetic fields and voltages than those derivable from equations (1) and (2). The first magnetron had an efficiency in the neighbourhood of 10%, and the pulsed power output of some of the early valves was approximately 50 kw. Under the new operating conditions efficiencies of more than twice the above figure were found. Later (1942) it became clear that the high operating conditions were simply an approach to the π-mode.

The mention of power and operating conditions brings up again a feature with which we were concerned in the earliest stages—that of high current or peak-current. From oxide-coated cathodes it is possible to get peak-currents of the order of 5 amp./cm.² of cathode surface, and from this fact it is seen that the limit of moderate beam current which obtains in the klystron is overcome.

Any discussion of cathodes would be incomplete without mention of secondary emission. From various pieces of evidence it became clear that electrons striking the cathode with finite velocity on completion of an orbit produced secondary electrons. This led to an interesting experiment in which a magnetron was made with a simple aluminium cathode entirely unheated. A small quantity of hydrogen was admitted to the valve from a palladium tube and was sufficient to start the build-up of current, and such a magnetron worked normally apart from eventual clean-up of the gas. This experiment showed for the first time the possibility of cold or even water-cooled cathode magnetrons, and foretold the ultimate, but still unrealized, development of such valves for c.w. purposes.

The next stage of development in the magnetron has come to be known as " strapping ". In this development I have no personal share, but the story of the magnetron would be grossly inadequate without it. Essentially this is the work of Dr. Sayers at Birmingham.

So far we have considered the magnetron as radiating only one frequency, and, indeed, this can be so. The resonator system has an infinite number of resonant modes with wave-lengths extending from some maximum value to lower and still lower values. The symmetry of the system causes the modes to be bunched together into groups of different frequencies, and each group by analogy with ordinary spectroscopy can be called a multiplet. In general each multiplet will contain $N-1$ modes if there are N resonators. Most of the modes are degenerate, i.e. more than one mode is found for a given wave-length. All

resonant wave-lengths but one (if N is even), or all (if N is odd) are degenerate. Thus in a 10-hole magnetron there are five possible wave-lengths in a multiplet $[1+(N/2-1)]$; one of these corresponds to a single mode, and each of the others to two modes, giving a total of $1+2(N/2-1)=N-1$ modes or $1+2\times4=9$. The fundamental fact of the existence of different modes in a magnetron resonator system showed itself in the phenomenon known as mode-jumping. For a given set of operating conditions a valve could suddenly jump from one of the possible modes to another nearby one. This had obvious disadvantages, and I should like to describe briefly how this was overcome by Sayers at Birmingham.

By connecting alternate segments together by short copper wires the circuit constants of the resonator system are altered, and the general effect can briefly be described as a greater separation of the modes of the multiplet. In nine- or ten-centimetre magnetrons this enables what is called the π-mode of operation, with a phase difference of π between adjacent segments, to be achieved. Examination of the wave pattern inside a cold block in these circumstances shows that the standing wave pattern is very regular in amplitude. The patterns for other modes are not suppressed but are irregular and often only weakly excited. The CV64, a 9–10-cm. strapped magnetron designed at Birmingham and greatly used by the Services, is illustrated in figure 4. The possibility of producing magnetrons having efficiencies of about 50% caused most of the British magnetron effort to be directed towards the design of strapped valves for particular Service applications at both 10 cm. and 3 cm. For the former wave-length no great difficulties were experienced. At first, considerable difficulties existed at 3 cm., but these cannot be detailed here; they were eventually overcome by adopting the American technique of scaling down a 10-cm. valve.

The maximum peak power now obtainable from cavity magnetrons in the 9–10-cm. region is of the order of a megawatt. Two views of such a valve are shown in figures 5 and 6. This is a great advance on the position existing in 1940, and is an example of what can be done by cooperation between the various organizations (universities, firms, establishments) responsible.

All resonator magnetrons operate above cut-off, that is to say, at a value of magnetic field for which there would be no anode current under steady conditions. There is a space-charge between anode and cathode, and the steady states without any high-frequency field on the anode have been analysed in detail by Hartree and Stoner.

The theoretical approach to the problem of the interaction of the electron stream and the high-frequency wave which travels round the anode is extremely complicated; that many of the problems concerning the operation of cavity magnetrons have been solved is in large part due to the efforts of Hartree, Slater, Stoner and their collaborators. Perhaps the most important early theoretical work was carried out by the Slater group. First, calculations were made utilizing field theory to solve the problem of the possible number of frequencies of a multiple resonator system. Secondly, the significance of the space-charge in the magnetron was realized and investigations based on the assumption of a linear field distribution between anode and cathode were made. From a knowledge of known operating conditions it was found possible to calculate the mode of operation of a given valve.

Later theoretical developments were in large part due to English workers, chiefly Hartree and Stoner. The team working under Hartree considered the problem of magnetron space charge more thoroughly and made calculations

based on a field distribution between anode and cathode which was a good deal nearer reality. This work led to the formulation of the now well-known threshold criterion. It was shown that exact synchronism between the outer edge of the space-charge cloud and the rotating wave was not required, but that operation of the valve, once started, would continue over a range of angular velocities of the cloud. The lower limit of angular velocity depended on the amplitude of oscillation ; it was possible to find a valve for which electrons would reach the anode for vanishingly small segment amplitudes for any practical mode-number n. Thus an anode voltage was determined below which no current could reach the anode for a given magnetic field, and this voltage has been called the threshold voltage. The threshold relation may be expressed in equation form by

$$\frac{V_T{}^2}{10100a^2} = \frac{2H}{n} \cdot \frac{1 - b^2/a^2}{21400} - \frac{1}{n^2},$$

where V is the threshold voltage in kilovolts, and n is the effective n number of the mode.

By the middle of 1942 the calculations, which have been very briefly outlined, had given considerable insight into the operation of valves in the 9-cm. to 12-cm. wave-length region. Large discrepancies, however, still existed between the calculated and observed operating conditions for 3-cm. magnetrons. In May 1942, Sayers and Sixsmith, by phase measurements on 3-cm. unstrapped valves with 12 to 16 segments, showed that such valves were not operating in the π-mode; furthermore, it was shown that the direction of the rotating anode-potential wave was opposite to that of the rotating space-charge cloud. This meant that a lower space-charge angular velocity was required for proper interaction for modes other than the π-mode; the lower space-charge velocities corresponded much more closely with the observed operating conditions. The Hartree group worked ou the required modification to the earlier resonance criterion. Under the new conditions it was shown that for electrons to reach the anode at magnetic fields well above cut-off value, and for vanishingly small high-frequency amplitude a certain degree of synchronism is required between the rotating space charge and one of the spatial components of the rotating electric field in the anode-cathode space.

The theoretical work on the magnetron which has been referred to has not yet been published, and it is extremely difficult to cover it adequately in such a short space. It is to be hoped that those directly concerned will take steps to give a comprehensive account in the near future.

REFERENCES

HANSEN, W. W., 1938. *J. Appl. Phys.*, 9, 654.
HANSEN, W. W. and RICHTMYER, R. D., 1939 *J. Appl. Phys.*, 10, 189.
JENKINS, H. G., 1945. *G.E.C. Journ.*, 13, 185.
MACDONALD, H. M., 1902. *Electric Waves* (Cambridge : University Press).
McKEAG, A. H. and RANDALL, J. T., 1936. British Patent 480356.
RANDALL, J. T. and WILKINS, M. H. F., 1946. *Proc. Roy. Soc.*, A (in course of publication).
RAYLEIGH, Lord, 1897. *Phil. Mag.*, 43, 125. (See *Collected Papers*, vol. IV).
VARIAN, R. H. and VARIAN, S. F., 1939. *J. Appl. Phys.*, 10, 321.
WILKINS, M. H. F. and GARLICK, G. F. J., 1945. *Proc. Roy. Soc.*, A, 184, 408.

AN APPARATUS FOR ACCURATE MEASUREMENT OF THE ACOUSTIC IMPEDANCE OF SOUND-ABSORBING MATERIALS

By R. A. SCOTT,
Manchester

ABSTRACT. This paper describes the principles and the construction of an apparatus for the precise measurement of the acoustic impedance, at normal incidence, of sound-absorbing materials by a stationary-wave method for frequencies in the range 100 to 5000 c/s. From measurements made on a sample of material $1\frac{3}{4}$ inches in diameter, backed by a substantially rigid wall, the magnitudes of the resistive and reactive components of the impedance may be calculated with an accuracy of about one per cent.

The theory of the method is discussed with particular reference to the influence thereon of the attenuation of sound associated with the walls of the tube in which the standing wave is formed, and it is shown that in addition to the correction which must be provided to the elementary expression for the " standing-wave ratio ", the finite attenuation in the tube leads to an additional correction to the expression for the distance from the sample of the first minimum of pressure. Attention is drawn to convenient methods which facilitate calculation of the results, and typical results of measurements of the impedance of a sample of porous sound-absorbing material are shown.

---

§1. INTRODUCTION

IN recent years interest has been steadily growing in the recognition of the acoustic impedance, rather than the absorption cofficient, as the chief characteristic of the performance of acoustical materials. The advantages which follow from a knowledge of the impedance are twofold. Firstly, modern theories relating to the absorption of sound both in enclosures (Morse, 1939a) and in ducts (Morse, 1939b) are formulated in terms of the impedance of the walls for normal incidence. Secondly, a knowledge of the separate resistive and reactive components of impedance of an absorbing material simplifies the task of relating the deficiencies in the absorbing power to the physical structure of the material.

Various methods for the measurement of acoustic impedance have been devised, and are referred to in a recent paper by Beranek (1940). There is a paucity of published results by such methods, and until the publication of Beranek's paper, no substantial accumulation of results of adequate accuracy had appeared. Beranek attributes this condition to the difficulties inherent n the methods of measurement previously in use, and proceeds to describe an apparatus of his own design which he considers to be capable of producing results of the required accuracy.

In the present paper, a modification of the apparatus originally used by H. O. Taylor is described. This apparatus is capable of an accuracy of the same order as that of Beranek and possesses, in the author's opinion, advantages of

simplicity in design and in theory of operation. Measurements of acoustic impedance can be made with this apparatus at frequencies in the range 100 to 5000 c/s.

§2. THE METHOD AND ITS THEORY OF OPERATION

The principle of the method is the same as that used by Paris, which in turn .s based on an extension of that devised by Tuma, and later used by H. O. Taylor for the measurement of the absorption coefficient. The sample of material closes one end of a uniform tube, and plane sound waves are set up in the tube by means of a source of sound at the other end. The stationary-wave pattern which is formed in the tube is explored by means of a microphone, and measurements are made of the ratio of the maximum to the minimum pressure in the system and of the distance of the first minimum from the face of the sample. From these two quantities the two components of the acoustic impedance of the material are obtained.

The derivation of the algebraic expressions which relate the resistive (R) and reactive (X) components of the acoustic impedance of the sample to the ratio of maximum (p_{max}) to minimum (p_{min}) oscillatory pressure in the stationary wave and to the distance (d_N) of the Nth minimum from the face of the sample is given in standard text-books on sound * for the ideal case of negligible dissipation in the stationary wave. Thus it can be shown that the acoustic impedance Z of the sample is given by

$$Z = R + jX = \rho c \tanh(\gamma + j\delta), \qquad \ldots\ldots(1)$$

where ρ is the density and c is the velocity of sound in the air. The quantities γ and δ are obtained from the measured values of $|p_{max}|/|p_{min}|$ and of d_N by the relations

$$|p_{max}|/|p_{min}| = \coth\gamma, \qquad \ldots\ldots(2)$$

$$d_N/\lambda = \tfrac{1}{2}N - \delta/2\pi, \qquad \ldots\ldots(3)$$

where λ is the wave-length of sound in the tube.

The assumption of negligible attenuation in the tube is not justifiable where precise measurements (i.e. to about 1 % in R and X) are required. The principal effect of finite attenuation in the tube is to cause the maxima and minima to alter progressively in magnitude with distance from the sample. An example of this is shown in figure 1, which gives a graph of oscillatory pressure against distance from the sample for a specimen of rock-wool 1″ deep at 1500 c/s., backed by a massive piston. The theory of this effect has been formulated by Davis and Evans (1930) and by Hall (1939), who show that to a first approximation the maxima can be regarded as substantially unaffected, whilst the minima increase linearly as the distance from the sample increases.

The finite attenuation in the tube leads also to a shift in position of the minima. The existence of this shift must have been clear to Davis and Evans, but its magnitude does not yet appear to have been discussed in detail. Neglect of the

---

* Morse (1936). Note that Morse describes the time dependence by $\exp(-j\omega t)$, whereas here we use the alternative $\exp(j\omega t)$. The difference results in a corresponding change of sign in the phase angle.

shift can cause appreciable error in precise measurements of impedance, and an outline of the effect will, therefore, be given.

Assume that the acoustic disturbance in the tube may be represented by means of two oppositely and axially directed trains of plane waves of frequency $\omega/2\pi$, wave-length constant $\beta\ (\equiv 2\pi/\lambda)$ and attenuation constant α. The expression for the oscillatory pressure p_x at a section of the tube distance x from the face of the sample is

$$p_x = A \sinh\{\alpha x + \gamma + j(\beta x + \delta)\},$$

whence the amplitude of the oscillatory pressure is given by

$$p_x = B\left[\cosh 2(\alpha x + \gamma) - \cos 2(\beta x + \delta)\right]^{\frac{1}{2}}, \qquad \ldots\ldots(4)$$
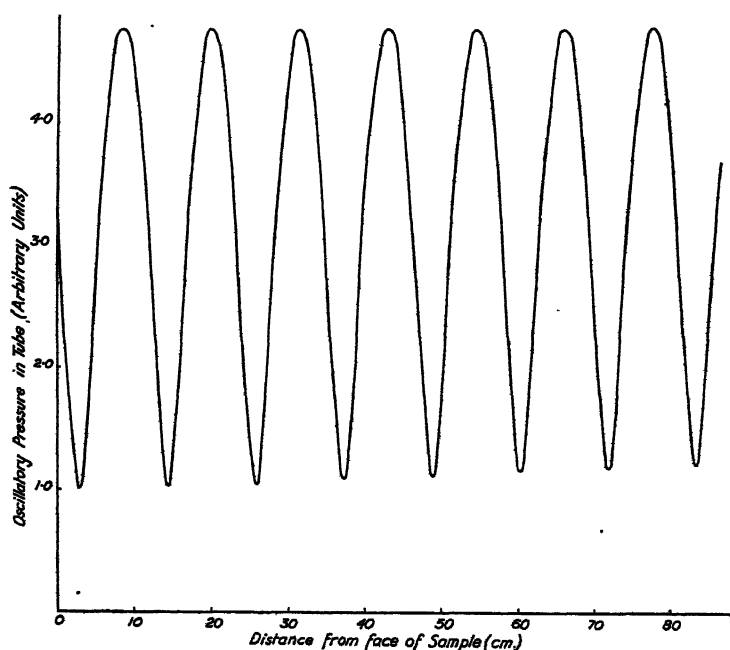
Figure 1. Typical graph of pressure against distance for sample.

where B is a constant independent of x, and where γ and δ are quantities which are associated with the conditions of reflection at the sample, and which satisfy equation (1); thus $e^{-2\gamma}$ is the ratio of the amplitudes of reflected to incident wave at the sample, and 2δ is the phase change accompanying reflection.

Stationary values of p_x occur when

$$\alpha \sinh 2(\alpha x + \gamma) + \beta \sin 2(\beta x + \delta) = 0, \qquad \ldots\ldots(5)$$

i.e.

$$\sin 2(\beta x + \delta) = -\frac{\alpha}{\beta}\sinh 2(\alpha x + \gamma).$$

If α/β is much less than π, the solutions of this equation can be written

$$2(\beta x + \delta) = 2N\pi - \frac{\alpha}{\beta}\sinh 2(\alpha x + \gamma)$$

and
$$2(\beta x + \delta) = (2M+1)\pi + \frac{\alpha}{\beta}\sinh 2(\alpha x + \gamma),$$

whence, for p_x a minimum, the distance d_N to the sample is given by

$$d_N = \tfrac{1}{2}N\lambda - \frac{\delta\lambda}{2\pi} - \frac{\alpha}{2\beta^2}\sinh 2(\alpha x + \gamma), \qquad \ldots\ldots(6a)$$

and for p_x a maximum, the distance D_M to the sample is given by

$$D_M = \frac{2M+1}{4}\lambda - \frac{\delta\lambda}{2\pi} + \frac{\alpha}{2\beta^2}\sinh 2(\alpha x + \gamma). \qquad \ldots\ldots(6b)$$

In the measurement of impedance, particular interest centres on the position of the first minimum. Rearranging (6 a),

$$d_N/\lambda = \tfrac{1}{2}N - \frac{\delta}{2\pi} - \frac{\alpha}{2\beta^2\lambda}\sinh 2(\alpha x + \gamma). \qquad \ldots\ldots(7)$$

Comparison of (7) with (3) shows that the effect of finite attenuation in the tube is to reduce the distance between the sample and the first minimum by an amount

$$\frac{\alpha}{2\beta^2}\sinh 2(\alpha x + \gamma).$$

The effect of attenuation on the ratio $|p_{max}|/|p_{min}|$ may be found by substituting from (6 a) and (6 b) in the expression for p_x. Thus for the ratio of the Mth maximum to the Nth minimum

$$|p_{max}|/|p_{min}| = \left[\frac{2\cosh^2(\alpha D_M + \gamma) - \dfrac{\alpha^2}{2\beta^2}\sinh^2 2(\alpha D_M + \gamma)}{2\sinh^2(\alpha d_N + \gamma) + \dfrac{\alpha^2}{2\beta^2}\sinh^2 2(\alpha d_N + \gamma)}\right]^{\frac{1}{2}}.$$

The importance of the effect of attenuation in the tube depends on the magnitude of the attenuation coefficient. For tubes with smooth walls and of diameters greater than six inches, the values of α are very small. Experiments on the apparatus which is to be described later show that for frequencies up to 3000 c/s. the attenuation constant for a tube of diameter 4·45 cm. is given by $1·37 \times 10^{-5}\sqrt{f}$ for room temperatures of 22° c. This figure corresponds to air of low humidity, and some modification may well be expected for damp air (Knudsen, 1933). The above figure is in good agreement with the theoretical attenuation constant deduced from the well-known Helmholtz-Kirchhoff analysis of the passage of sound waves along smooth-walled, circular pipes (viz. $1·33 \times 10^{-5}\sqrt{f}$ at 23° c.).

Where γ is small compared with αd_1, the shift, which is a maximum at low frequency, is less than 0·1 % of d_1. As γ approaches unity, however, the correction becomes

$$\frac{\alpha}{2\beta^2}\sinh 2\gamma,$$

and is appreciable at low frequencies. Figure 2 shows a plot of $\sinh 2\gamma$ against the ratio $|p_{max}|/|p_{min}|$ expressed in terms of $L = 20\log_{10}|p_{max}|/|p_{min}|$, and included on the graph is a table of $\alpha/2\beta^2$ for various frequencies in the range of the apparatus. For a stationary wave with $|p_{max}|/|p_{min}| = 2$, i.e. $L = 6·0$, $\sinh 2\gamma = 1·34$ and the

correction to d_N becomes 2·7 mm. at 100 c/s. For a tube of smaller diameter this correction would become more important.

For the values of α and β appropriate to the present apparatus, the ratio of maximum to minimum pressure differs negligibly from

$$\cosh(\alpha D_M + \gamma)/\sinh(\alpha d_N + \gamma).$$

Successive maxima lie on a cosh curve and minima lie on a sinh curve. The maxima vary very little with distance, whilst the minima vary approximately
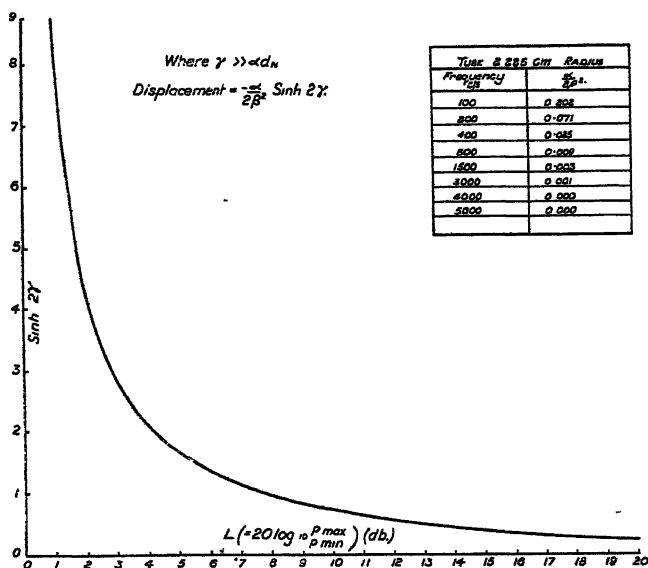
Figure 2. Correction to position of minimum, for effect of attenuation in the tube.

linearly with d_N, except where γ is large, in which case the total variation of minima with distance is correspondingly small. Account may conveniently be taken of the effect of attenuation on the ratio of maximum to minimum by plotting $|p_{max}|$ and $|p_{min}|$ against distance from the sample and extrapolating the envelopes to the position of the sample. The ratio $|p_{max}|/|p_{min}|$ extrapolated in this fashion then satisfies the equation (2), which is deduced by neglecting attenuation.

§3. DESCRIPTION OF APPARATUS

For most practical problems in which sound-absorbing materials are used, it is necessary to know the magnitudes of the components of the acoustic impedance to within 5%. A better accuracy than 5% is required where measurements of impedance are to be used in the discussion of the detailed mechanism by which sound is absorbed in the material. It is, therefore, desirable that the measuring apparatus should be capable of providing results with an accuracy of 1% in both resistive and reactive components for frequencies in the range of 100 to 5000 c/s. At lower frequencies than 100 c/s., available materials have such poor absorption that accurate measurements of impedance are at present of little practical interest.

Where determination of the normal sound-absorption coefficient to within 0·01 is concerned, measurements present no great difficulty, since the ratio

$|p_{max}|/|p_{min}|$ alone is involved. For measurements of impedance a knowledge of the position of the minima is also necessary, and here the requirements are stringent. The impedance of a layer of almost any material used for absorption of sound possesses a large negative reactance term for frequencies low in the acoustic range. The reactance is associated with the finite depth of the absorbing material, and is due to the high elasticity of the air trapped in the layer. Thus $R/\rho c = 1 \cdot 01$, $X/\rho c = -16 \cdot 7$ for a $1''$ thick layer of rock-wool at 100 c/s. In this case d_1 is about 83 cm., λ about 345 cm. and d_1 must be measured to within $0 \cdot 3$ mm. in order to give R and X with an accuracy of 1%. If such precise measurement is to be significant, the following conditions must be satisfied :

(i) the bore of the tube must be uniform and the walls smooth and substantially rigid;

(ii) the sound field in the tube must consist of plane, uniform, axially directed waves;

(iii) the face of the specimen of material must be substantially plane and mounted normal to the axis of the tube;
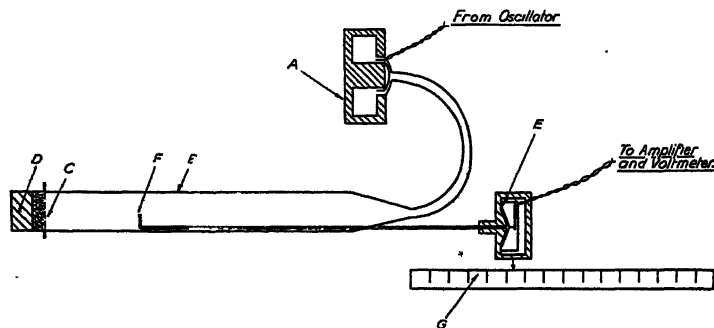
Figure 3. General arrangement of measuring apparatus

(iv) the microphone used for exploration of the sound field must not itself appreciably affect the field and must be sensitive and stable;

(v) the measurement of position of microphone orifice must be accurate to about 0·1 mm.;

(vi) the pressure measured with the microphone must arise only from sound of single frequency.

The apparatus was designed and constructed so as to satisfy these requirements. The conditions under which plane, axially directed waves may be set up in a tube are well known. For a circular tube Rayleigh has shown theoretically that serious departures from uniform plane waves will not occur provided that the diameter of the tube is less than $0 \cdot 58 \lambda$, where λ is the wave-length of the sound of highest frequency used in the tube. Experiment shows that in practice the field is substantially uniform until the diameter of the tube reaches nearly twice this value. Thus, taking Rayleigh's limit, if measurements are required at frequencies as high as 5000 c/s., the diameter of the tube should not exceed 4·0 cm. A diagram of the apparatus is shown in figure 3. The tube (B) has an internal diameter of 4·45 cm. and a smooth brass wall $\frac{1}{8}''$ thick. The

maximum variation in cross-sectional area along its length amounts to less than 0·1 %. The specimen of absorbent material (C) is mounted in a detachable specimen tube which is closed at the rear end by a massive (1000 gm.) close-fitting brass piston (D). The positions of the piston and specimen are adjusted so that the front face of the specimen lies accurately in the plane (H) of the flange of the specimen tube. The sound enters the tube through a conical end-piece (J) from a moving-coil loud-speaker unit (A) (Vitavox: type N), which in turn is supplied with electrical power from a valve oscillator of conventional type. The sound field in the tube is explored with a moving probe-microphone (E), the position of which, along the axis of the tube, is measured by means of a cursor sliding over an accurately engraved steel metre scale (Chesterman: No. 1367) (G) divided in millimetres.

The mode of construction of the probe-microphone is important. The microphone consists essentially of a long tube of stainless steel of internal diameter 0·241 cm., external diameter 0·316 cm. and length 104 cm. One end of the tube is connected to a small cavity, of volume about 2 cm³, in front of the conical diaphragm of an otherwise hermetically sealed rochelle-salt, piezo-electric microphone. The passage of sound waves to the microphone by any path other than along the tube is prevented by the presence of the massive ($\frac{1}{2}''$ thick) shell of the microphone casing, which is made from a steel bar of diameter 3″ bored out so as to take the microphone unit. The other end (F) of the probe-tube is open and communicates with the sound field. The response of the microphone at any one frequency is proportional to the oscillatory pressure at the mouth of the probe-tube, and since the impedance presented by the mouth of the probe-tube is high, the acoustic pressure at the mouth of the probe is effectively that existing at the same position in the sound field when the microphone is withdrawn. The modification of the sound field caused by obstruction due to the probe-tube is negligible, since the area of cross-section of the probe-tube is less than $\frac{1}{2}$ % of that of the main tube. The construction of the apparatus is shown in more detail in figures 4 and 5. Figure 4 is a side view of the trolley (with indicating meter removed) and of the driving end of the tube. Figure 5 is an end view of the trolley, looking from the direction of the tube.

Vibration which exists in the wall of the probe is prevented from reaching the microphone by the interposition of a sleeve (K) of cycle valve-rubber between the two components. Tests were made of the response of the microphone in the normal condition and with the orifice of the probe-tube closed with a short length of brass rod. A typical test at 1000 c/s. showed that the response of the microphone was reduced at least by 50 db. when the orifice was closed. The microphone has proved in practice to be stable and sensitive, and no difficulty is experience in determining the acoustic pressure at maxima or minima in the tube to within 0·1 db. for any frequency in the range from 100 c/s. to 5000 c/s.

The probe-tube passes into the main tube through a long hole bored in the conical end-piece, and for the most part of its length rests on the bottom of the main tube. The probe-tube, rather than the microphone casing, is clamped rigidly to the saddle of the microphone trolley (L), whilst the microphone casing is resiliently supported above the trolley by means of the sponge-rubber (M) and the ring (N). The scale cursor (O) is mounted close to the fixing point

of the probe, and in consequence the scale readings give accurately the successive positions of the probe orifice.

The microphone trolley is mounted on four wheels: two grooved to fit one rail of the angle-brass track (Q in figure 5) and the other two ungrooved. Coarse
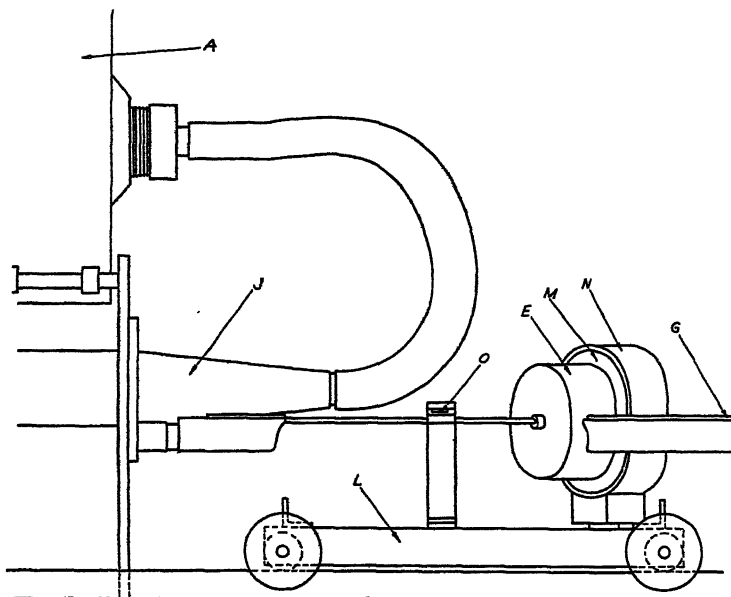
Figure 4. Detail of driving unit and microphone trolley.

adjustment of position of the microphone is made by pushing the trolley bodily along the track, and fine adjustment by means of the knurled knobs (R), which form extensions to the near-side wheels. A four- rather than three-wheeled construction for the trolley was used to provide greater stability of the trolley.
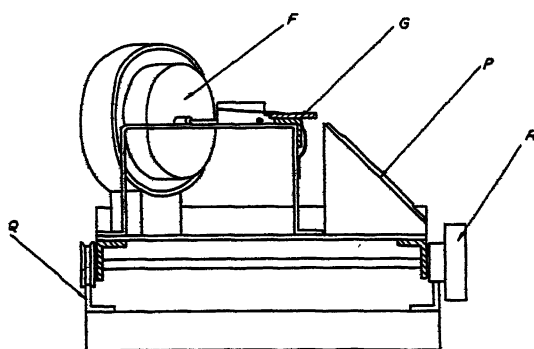
Figure 5. End view of microphone trolley.

The frame of the trolley was made to be sufficiently flexible for each of the wheels to remain in good contact with the rails.

The output voltage from the microphone is amplified by means of a multi-stage valve-amplifier, and is then passed through an electrical analyser, tuned to the working frequency, to an indicating meter (P in figure 5) of the moving-coil,

rectifier-type (2½″) which, for convenience, is mounted on the trolley. The amplifier effectively excludes such harmonics of the fundamental frequency as may exist in the sound field.

In the plate facing p. 264, (*a*) shows a general view of the apparatus, and (*b*) gives a more detailed view of the driving end of the tube and of the microphone trolley.

Typical operation of the apparatus is as follows :—A sample of acoustical material is cut by means of a circular pastry-cutter so as to fit accurately into the specimen tube. The front face of the material is adjusted so as to lie accurately in the plane of the face of the flange of the specimen tube, and the brass piston (D) is moved so as just to touch the rear face of the specimen. The sound field of appropriate frequency is set up in the tube; there is, of course, no necessity to "tune" the tube since the form of the sound-field, looked at from the "sample" end of the tube, depends only on the nature of the reflection at the sample. The reading of the indicating meter, which is scaled in db., now corresponds to the pressure at the end of the microphone tube. The scale (G) and cursor have previously been set up so as to give a reading of zero when the axis of the upturned end portion of the probe-tube is exactly in the plane of the face of the flange to which the specimen tube attaches. Corresponding readings of indicating meter and of scale give the acoustic pressure in terms of the distance from the face of the sample. The trolley is moved smoothly away from the specimen, and several successive readings are taken of pressure at maxima and minima. From these measurements the ratio at the sample of maximum and minimum pressure is obtained, if necessary by extrapolation of a graph of the ratio, in linear units.

Positions of minima relative to the sample are now found, and are conveniently determined from the mean of two readings, equally spaced about the minimum, each corresponding to the meter pointer passing through an arbitrarily chosen value just in excess of the minimum. The curve of meter reading against distance shows a high degree of symmetry in the region of the minimum. When the number of minima in the tube allows, the positions of eight minima (four at the specimen end and four at the distant end of the tube) are measured. These data allow not only a convenient check on each other, but also make it possible to determine directly the wave-length of the sound, which enters into the calculation.

Thus for a sample of rock-wool (stillite, 5 lb./cu. ft.) 1″ thick at 1500 c/s., successive maxima and minima are shown in the following table. The maximum remains closely constant at 13·6 db. (the reference level is arbitrary, but the same throughout the measurements). The acoustic pressure is plotted directly against distance from the sample in figure 1, and extrapolation back to the sample shows that, with no dissipation in the tube, the minimum pressure would have a value of 0·99. The ratio $|p_{max}|/|p_{min}|$ of maximum to minimum pressure is, therefore, 4·83, i.e. 13·7 db.

The positions of eight minima are also shown in the table. By pairing of readings, four values for 2λ are obtained and lead to a value of λ of 23·00 cm. and to a corresponding velocity of sound of 345·0 m./sec. The position of the first minimum (d_1) can be obtained directly from the first reading, viz. 2·895 cm., from the sample. This value can be checked by means of the other figures

Position of minima of pressure and magnitude of maxima and minima
for a typical sample of rock-wool

Stillite, 5·0 lb./ft³, 1″ depth

Order N	Intensity $\|p_{max}\|/\|p_{min}\|$ (db.)	Distance from sample d_N(cm.)	2λ from preceding column (cm.)	$d_1 = \left(d_N - (N-1)\frac{\lambda}{2}\right)$ (cm.)
1		0·0	2·895		2·895
	13·6				
2		0·2	14·39		2·89
	13·6				
3		0·5	25·89		2·89
	13·6				
4		0·75	37·39		2·89
	13·65		———		
5		1·0	48·89	45·995	2·89
	13·65				
6		1·25	60·395	46·005	2·895
	13·65				
7		1·45	71·895	46·005	2·895
	13·7				
8		1·65	83·395	46·005	2·895

Mean 46·00

since the successive readings d_N should differ from the first reading only by some known multiple $(N-1)$ of the measured half-wave-length. Thus the last column of figures in the table is found by subtracting from the measured values of d_N the appropriate multiple of the half-wave-length, and it is apparent that the agreement in the resulting values of d_1 is excellent.

Using the above data,

$$\|p_{max}\|/\|p_{min}\| = 4\cdot 83 \text{ or } 13\cdot 7 \text{ db.},$$
$$d_1 = 2\cdot 895 \text{ cm., i.e. } d_1/\lambda = 0\cdot 1259.$$

The correction to d_1 as calculated from (7) is 0·001 cm. The correction is not significant at this frequency. The corresponding values of $R/\rho c$ and of $X/\rho c$ are

$$R/\rho c = \quad 0\cdot 399,$$
$$X/\rho c = \ -0\cdot 927.$$

The length (86·5 cm.) of the tube used in the present apparatus is such that two or more minima can be located only for frequencies of 300 c/s. and above. Direct measurement of λ from the interval between minima is therefore impossible below 300 c/s., and λ must be found by other methods. Careful measurement of the wave-length, and hence of the velocity of sound, in the tube in the range of frequency 400 to 4000 c/s. has shown that the velocity C' follows closely the Helmholtz-Kirchhoff law,

$$C' = C\left(1 - \frac{\sqrt{v'}}{2r\sqrt{\pi}}\cdot\frac{1}{\sqrt{f}}\right),$$

where C is the free-space velocity, r is the radius of the tube, f the frequency and v' a function of the kinematic viscosity and diffusivity of the air (Rayleigh) equal numerically, for r in centimetres, to 0·577 at 23° c.

The usual procedure adopted with the apparatus is to measure C' accurately at one frequency (3000 c/s., at which many minima are available) and to apply a small correction calculated from the Helmholtz-Kirchhoff relation, at other frequencies. By this means the wave-length in the tube can be determined for frequencies below 300 c/s. and measurements of acoustic impedance can be made at 100 c/s., at which frequency the tube-length is only one-quarter of the wave-length.

§4. REDUCTION OF RESULTS

Apart from the corrections for attenuation, the values of $|p_{max}|/|p_{min}|$ and of d_1/λ are provided directly by the apparatus, and there remains the process of calculating the real and imaginary components of the impedance from them. This problem reduces to the solution of equations (1), (2) and (3). These equations can be solved algebraically, and thus it can be shown that

$$R/\rho c = \frac{2r}{1+r^2+(1-r^2)\cos 4\pi d_1/\lambda},$$

$$X/\rho c = \frac{-(1-r^2)\sin 4\pi d_1/\lambda}{1+r^2+(1-r^2)\cos 4\pi d_1/\lambda},$$

where
$$r = \tanh\gamma = \frac{|p_{min}|}{|p_{max}|}.$$

Numerical evaluation of $R/\rho c$ and $X/\rho c$ is , however, tedious, especially when results are to be worked out at several frequencies. It is convenient, therefore, to have graphical methods of solution available. In this respect the position is fortunate. H. J. Sabine (1942) has constructed charts with rectangular axes $R/\rho c$ and $X/\rho c$, on which are plotted families of curves of constant $|p_{max}|/|p_{min}|$ (in decibel units) and of constant d_1/λ. (He plots for D_1/D_2, where D_1 has the same significance as our d_1 and $D_2 = \frac{1}{2}\lambda$.) These charts are very useful for measurements at medium frequencies on materials of depth greater than 1 inch. For lower frequencies and small depths, e.g. where $R/\rho c$ may be of the order unity and $X/\rho c$ of the order -20, the range of curves is inadequate.

A second chart suitable for the calculation of $R/\rho c$ and $X/\rho c$ has been provided recently for the solution of a closely analogous problem in connection with radio wave-guides. Such a chart has been published by P. H. Smith (1939) and by Willis Jackson and L. G. H. Huxley (1944). In this chart, functions of $|p_{max}|/|p_{min}|$ and of d_1/λ are provided as polar co-ordinates, and curves of constant $R/\rho c$ and $X/\rho c$ are plotted thereon. The resulting families of curves corresponding to all values of $R/\rho c$ from zero to infinity and of all values of $X/\rho c$ from minus to plus infinity lie inside a circular boundary of finite radius. This type of chart is easily constructed from the data provided in the above references and is very convenient to use in practice.

Figure 6 shows the results of typical measurements of acoustical impedance obtained, in the manner described above, over the range 100 to 5000 c/s.,

of a 1″ thick specimen of rock-wool (stillite: 5 lb./ft³) backed by a substantially rigid piston.
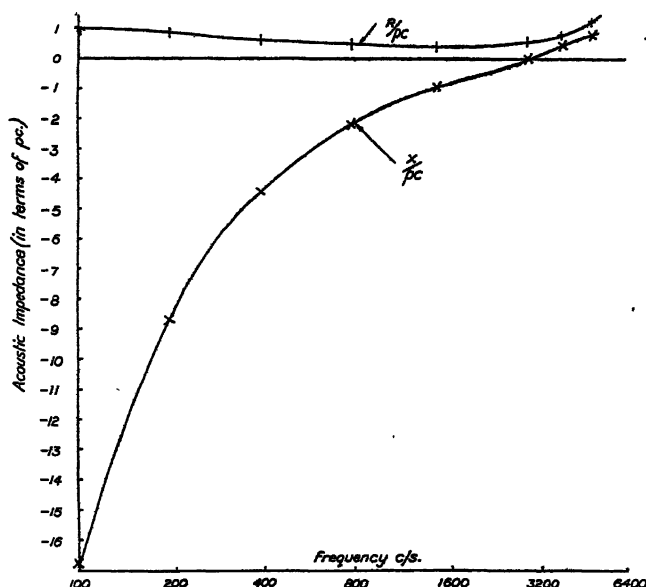
Figure 6. Typical measurements made with the apparatus of acoustic impedance of a sound-absorbing material.

ACKNOWLEDGMENTS

REFERENCES

BERANEK, L. L., 1940. *J. Acoust. Soc. Amer.*, 12, 3.
DAVIS, A. H. and EVANS, E., 1930. *Proc. Roy. Soc.*, A, 127, 89.
HALL, 1939. *J. Acoust. Soc. Amer.*, 11, 140.
JACKSON, WILLIS and HUXLEY, L. G. H., 1944. *J. Inst. Elect. Engrs.*, 91, 105.
KNUDSEN, 1933. *J. Acoust. Soc. Amer.*, 5, 122.
MORSE, P. M., 1936. *Vibration and Sound* (New York : McGraw-Hill Publishing Co.), p. 209.
MORSE, P. M., 1939 a. *J. Acoust. Soc. Amer.*, 11, 56.
MORSE, P. M., 1939 b. *J. Acoust. Soc. Amer.*, 11, 205.
RAYLEIGH, Lord, 1929. *Theory of Sound* (London : Macmillan), 2, 326.
SABINE, H. J., 1942. *J. Acoust. Soc. Amer.*, 14, 143.
SMITH, P. H., 1939. *Electronics*, 12, 29.
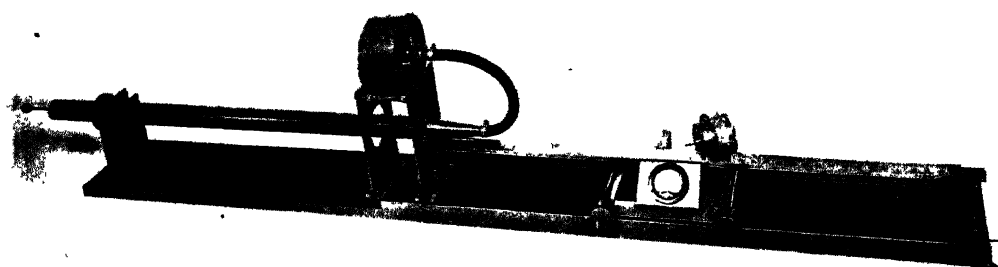
(a) General view of apparatus.
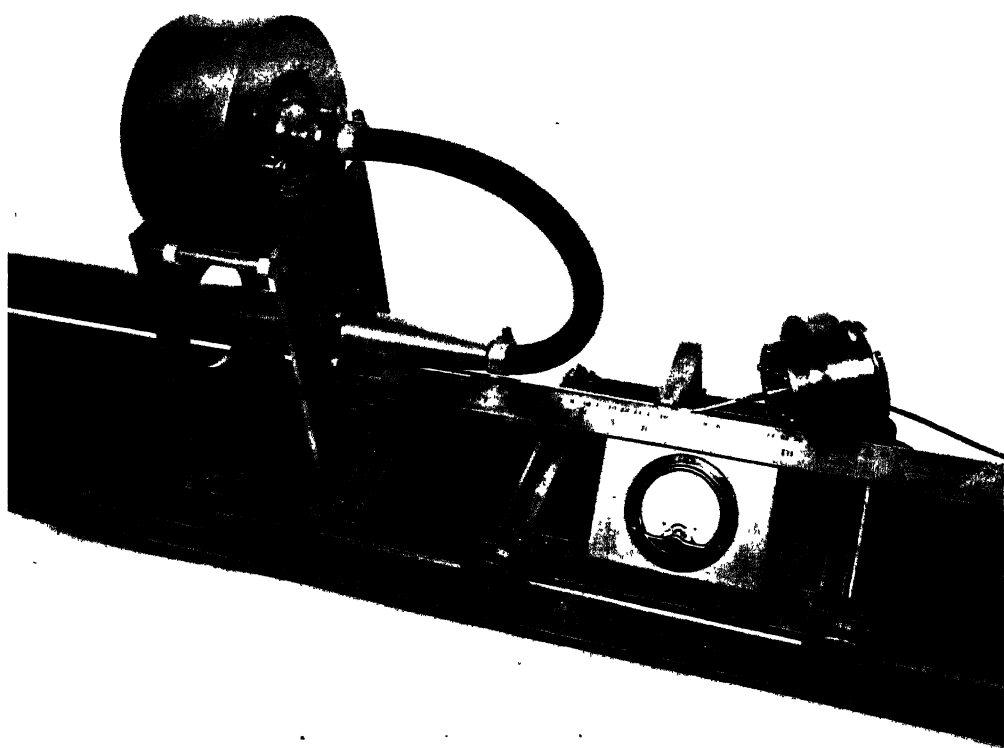
(b) View of driving end of apparatus and of microphone trolley.

General view of reflection experiments.

AN EXPERIMENTAL INVESTIGATION OF THE REFLECTION AND ABSORPTION OF RADIATION OF 9-CM. WAVE-LENGTH.

By L. H. FORD and R. OLIVER,

National Physical Laboratory, Teddington

ABSTRACT. The paper, which was circulated as a confidential report to the Radio Research Board in October 1944, describes measurements on the reflecting power of various surfaces at a wave-length of 9 cm. and angles of incidence ranging from 80° to 45°; these included level and uneven bare ground, vegetation-covered ground, tap water and a 4% salt solution approximating to sea water. Specular reflection was found to occur only from very level surfaces ; the absorptions of these surface media were measured, and from the combined measurements of reflection and absorption their electrical constants were derived.

Rough surfaces, either of bare ground or vegetation-covered ground, gave values of reflection coefficient in general agreement with the optical rule that regular reflection is only observed from an uneven surface if the product of the depth of the surface irregularities and the cosine of the angle of incidence is a small fraction of the wave-length. If this fraction exceeded $\frac{1}{8}\lambda$, the values of reflection coefficient measured were about 0·1.

The calculated values of reflection coefficient corresponding to dry soil, wet soil and sea water for angles of incidence varying from 0° to 85° are given in an appendix to the paper.

---

§ 1. INTRODUCTION

THE reflection of radio waves has been investigated experimentally at a wave-length of 9 cm., using a wide range of reflecting surfaces; these included bare ground, both level and in ridges, ground covered with vegetation, fresh water, and salt water approximating in composition to sea water. When specular reflection occurs, the electrical constants of the material forming the reflecting surface can generally be estimated from measurements of the reflection coefficient. At the high frequencies involved in this investigation, however, the reflection coefficient is determined almost entirely by the dielectric constant; consequently reflection experiments, while providing a reasonably accurate indication of the dielectric constant, only enable a very rough estimate to be made of the conductivity of the surface material. The absorption in the material, on the other hand, is then dependent mainly on the conductivity. Experiments were therefore made on the absorption, at the same wave-length, of the materials of the various reflecting surfaces, from which the conductivities could be determined.

Since both reflection and absorption experiments are needed to obtain the electrical constants, the sequence adopted in the presentation of this paper is to describe each of the experimental methods before giving any of the results obtained from either, rather than to divide the paper into two distinct sections. The results of both reflection and absorption measurements are then given together, for each type of surface in turn.

§2. EXPERIMENTAL METHODS

A. *Measurement of reflection coefficients*

The measurement of the coefficient of reflection for radio-waves is simplified at centimetre wave-lengths by the possibility of using aerial systems which give beams of radiation which are so narrow that, except at angles within a few degrees of normal or grazing incidence, the signals due to the direct and the reflected rays can be measured separately by suitably adjusting the aerial. The magnitude of the reflection coefficient is thus obtained from the ratio of two measurements of field strength; no indication of its phase angle is obtained.

A diagrammatic representation of the arrangement of the apparatus is shown in figure 1, while the plate shows a photograph of an experiment. The two towers were constructed of tubular scaffolding, and were 15 metres (50') apart, with platform levels of 1·7 m. (5·5'), 4·3 m. (14') and 6·7 m. (22') above ground. The
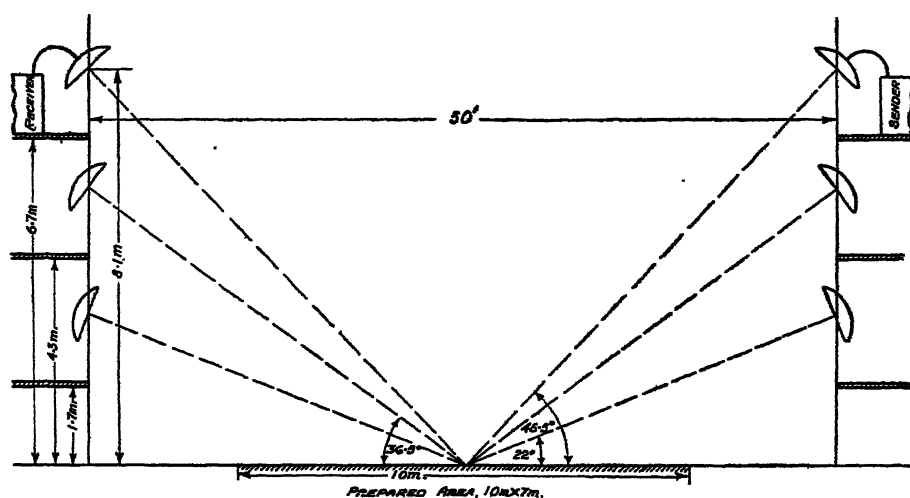
Figure 1. Diagrammatic representation of ground reflection experiments.

1·2-metre diameter aperture parabolic mirrors, each of which was used with a half-wave dipole aerial at the focus, were so mounted that the aerials were 1·4 metre above the platform level, thus giving angles between the reflected ray and ground of 22°, 36°·5 and 46°·5 (angles of incidence of 68°, 53°·5 and 43°·5). Between the two towers a symmetrically disposed area of ground with sides 7 × 10 metres was prepared to obtain the several types of reflecting surface used. For experiments on water, a National Fire Service canvas dam, giving a water surface of about 5 metres diameter, was placed symmetrically between the towers.

The sender consisted of a CV. 67 valve oscillator mounted directly on the back of one of the paraboloids, and a crystal monitor was used to ensure constancy of output during each set of observations. Two receivers were used: one (Saxton and Grace, 1946) was a bolometer mounted directly behind the receiving para-boloid, the other was a field-strength measuring set (Grace, 1946). When using the bolometer, reflection coefficients greater than about 0·2 could be obtained with an instrumental accuracy of about ± 0·01, from the square root of the ratio of the

values of the power received with a direct and with a reflected signal. When using the field-strength measuring set, the ratio of the direct and reflected received signals, in decibels, could be measured with an accuracy of about ± 0.5 db. if less than 6 db., and an accuracy of about ± 1 db. if greater than 6 db. Thus the bolometer was considerably more sensitive than the field-strength measuring set, except under conditions where the reflected signal was small.

The procedure followed at each level was to adjust the sending and receiving mirrors in turn until a maximum signal could be measured from the direct ray; next to readjust the mirrors for a maximum signal from the reflected ray, and finally to repeat the measurement of the direct signal to verify that the sender output had remained constant.

Apart from instrumental inaccuracies in measurement, there are a number of possible sources of error in the results obtained. Experiment showed that the polar diagrams of the mirrors were such that any signal due to the direct ray, when measuring the reflected ray, would be more than 40 db. down, which virtually eliminated one potential source of error. Experiment also showed that a distance of 15 metres between the mirrors was somewhat less than the minimum distance at which the transmission law holds, that the received signal varies inversely as the distance from the sender. A careful series of measurements was therefore made to determine the variation of received signal with distance from the sender over the range concerned, and the values thus obtained were used in place of the geometrical values when correcting the measurements of reflected signal for the path difference between the direct and reflected rays. A slight, but perceptible, interaction could, moreover, be detected between the receiver and sender mirrors, due to reflection from receiver to sender mirror when the two were directed at each other; calculation showed that this was unlikely to cause an error of more than 0.01 in the reflection coefficient. The agreement obtained between measurements at three different angles of incidence, and consequently three path differences, under conditions of specular reflection, makes it improbable that any of these sources of error was serious; consequently, the reflection coefficients measured with bolometer are believed to be accurate to ± 0.02. Had the distance between the towers been made great enough to reduce all the sources of error mentioned to negligible proportions, both the height of the towers and the area of ground requiring preparation would have become undesirably large.

To check the reliability of the method, a level platform of planks was set on the ground midway between the towers, and on it were placed six 1×2-metre iron sheets, making an area of 3×4 metres. The reflection coefficient of this was measured at angles of incidence of $53°.5$ and $43°.5$. The reflection coefficients obtained varied between 0.95 and 1.03, with a mean value of 0.98. The observations were made with the field-strength set, and the possible instrumental error was consequently ± 0.5 db. The values obtained, therefore, are in sufficiently good agreement with the expected value of unity to indicate that none of the possible sources of error in the experimental method was serious.

B. *Measurements of absorption*

The absorption of radiation by a screen of the reflecting medium used in the previous experiments was measured directly by interposing a filled container

between sender and receiver. The arrangement used is shown diagrammatically in figure 2. The sender was exactly as used in the reflection experiments, and was set up about 6 metres away from a table on which stood the test sample and the receiver. The latter consisted of a horn in which was a bolometer, forming the field-strength measuring device described by Saxton and Grace. A container consisting of a wooden trough was made having ends of 6 mm. plate glass, 46 cm. square, one of which was fixed and the other movable, thus forming a container in which a sample 46 cm. square and of any desired thickness up to 30 cm. could be tested. This container could be interposed between the sender and the receiving horn.

The experimental procedure was to measure the field strength at the horn when the transmission path from the sender was unobstructed, and then to repeat the measurement with the container filled with a chosen thickness of the substance under test, placed symmetrically to obscure the horn from the sender. Successive measurements were made with the container at various distances from 1 to 6 cm.
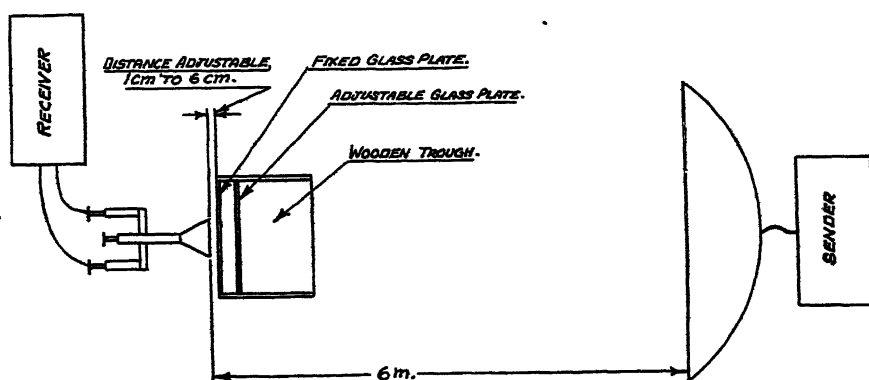
Figure 2. Diagrammatic representation of absorption experiments

from the aperture of the horn. In most cases considerable variations were found to be present in the received signal as the distance was varied, due, presumably, to reflection from the end of the container nearest the horn, and the mean value over a range of distance equal to half a wave-length was taken as the true transmitted signal. The thickness of the obstruction was then altered by adjusting the movable glass end-plate and refilling the container, and the experiment repeated. Thus the attenuation in a known thickness of the material under test could be obtained without reference to the properties of the container.

The method of measurement was subject to limitations; if the thickness of material under test were too small, multiple reflections between the two end plates would be a serious source of error, while if the thickness were too great the accuracy of measurement would be inadequate. With thicknesses of material such that the attenuation was between about 5 and 15 db., however, the measurements are believed to be accurate to ± 0.5 db.

§3. EXPERIMENTAL RESULTS: SPECULAR REFLECTION

A. *Reflection and absorption from bare ground*

An area 7×10 metres midway between the towers was cleared of turf, forked over to a depth of about 10 cm., carefully raked level, and finally smoothed by

dragging a plank across it. Preliminary experiments showed that the surface thus obtained gave specular reflection. This conclusion was based on the fact that a single value of dielectric constant, K, and conductivity, σ, could be chosen which gave calculated values of reflection coefficients which, at all angles and with both horizontal and vertical polarizations, agreed within the accuracy of experiment with the measured values. Curves connecting the reflection coefficient with K and σ, of the type described by McPetrie (1938), were drawn for each of the angles of incidence available to simplify the choice of the appropriate values of K and σ to fit any given set of test results. It quickly became apparent that the dielectric constant rather than the conductivity determined the reflection coefficient, and that changes in the wetness of the ground caused large variations in the value obtained. The decision was therefore made to concentrate on the two extreme conditions of humidity.

The penetration of the wave into the ground on reflection with centimetre wave-length radiation is only a few centimetres, which is much less than occurs at

(a) Soil.   (b) Water.
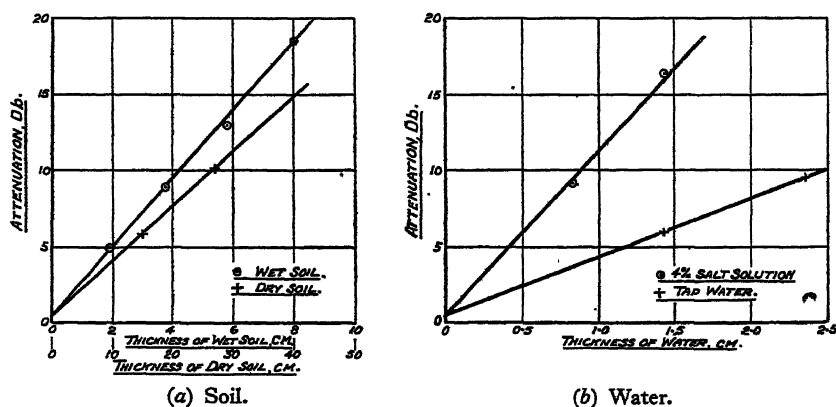
Figure 3. Attenuation of $\lambda = 9$ cm. radiation in various media.

the longer wave-lengths; consequently extreme conditions, particularly of dry ground, will be more often encountered with centimetre waves than with the longer wave-lengths.

An immediate consequence of the relative unimportance of the conductivity in determining the reflection coefficient is that the change in phase on reflection is about 180° for horizontally polarized waves at all angles of incidence; it is about 180° for vertically polarized waves at angles of incidence appreciably greater than the Brewster angle, and at angles of incidence appreciably less than that angle there is no appreciable phase change on reflection.

(a) *Dry ground.* A prolonged period during May 1944 in which no rain fell gave an opportunity to measure the ground reflection under conditions where the surface layer was as dry as is ever probable; no trace of moisture was apparent, and the top surface was almost as loose in consistency as sand.

The decision to measure the ground absorption was not taken until some time after the reflection tests had been made. Consequently, to obtain ground in as nearly as possible the same condition, a sample of surface soil was placed under cover and left for about the same period as had elapsed without rain before the reflection test. The results of the absorption tests on dry soil are shown in figure 3 (a), and indicate an attenuation of 36 db. per metre.

The most suitable values of K and σ to fit both reflection and absorption measurements were

$$K = 2, \qquad \sigma = 3 \times 10^8 \text{ e.s.u.}$$

The reflection coefficients calculated from these values, together with the measured values, are given in table 1.

Table 1. Reflection coëfficients for very dry ground, $\lambda = 9$ cm.

Angle of incidence (degrees)	Vertical polarization		Horizontal polarization	
	Calculated	Observed	Calculated	Observed
68	0·18	0·20	0·48	0·47
53·5	0·015	0·03	0·33	0·33
43·5	0·08	0·09	0·26	0·27

(b) *Wet ground.* The preliminary experiments showed that the value of K corresponding to the reflection coefficients usually encountered varied between 5 and 15, according to the degree of wetness of the ground. The opportunity was taken during a rainy spell in February to obtain a series of measurements immediately after a twelve-hour period of steady rain, when the ground was well saturated. The field-strength measuring set was used, so that the results obtained were of less accuracy than the others given in this report for conditions of specular reflection, all of which were taken using the bolometer. Accordingly, a further series of measurements were made in which, since the weather conditions were not suitable, the ground was kept wet artificially. For this purpose a spray branch of the type used to deal with electrical fires was used with a fire hose to water the site, and the equivalent of about four inches of rain was delivered at intervals over a period of 48 hours. A slight further watering was given about half an hour before each test. The resulting condition of the ground was such that it was saturated with water, but was not waterlogged, i.e. no puddles were present on the surface.

To obtain a similar sample of soil for absorption measurements the container was filled with soil and water, and allowed to stand until the surplus water had drained off. It was found that this method gave closely repeatable results, and the assumption that the sample in the container was then under the same conditions of humidity as the ground during the reflection experiments seemed justified. The results of the absorption tests on saturated soil are shown in figure 3 (*a*), and indicate an attenuation of 220 db. per metre.

The most suitable values of K and σ to fit both reflection and absorption measurements were

$$K = 24, \qquad \sigma = 6 \times 10^9 \text{ e.s.u.}$$

The reflection coefficients calculated from these values and the measured values are given in table 2; calculated values assuming $K = 25$, $\sigma = 0$, are also included to demonstrate the comparatively small contribution made by the conductivity to the reflection coefficient.

The ground surface on which the above experiments were made is composed of a fine sandy loam. Other types of ground might give appreciably different results.

The conductivities of both dry and wet soil were measured at a frequency of 1 Mc/s. ($\lambda = 300$ m.) by the Electricity Division, N.P.L., using a laboratory method. The values obtained were 3×10^5 and 8×10^7 e.s.u. for dry and wet soil respectively. These agree with the conductivities given by Smith-Rose (1933) for the soil at Teddington in extreme conditions of low and high humidity, and

Table 2. Reflection coefficients for saturated ground, $\lambda = 9$ cm.

| Angle of incidence (degrees) | Vertical polarization | | | | Horizontal polarization | | | |
| | Calculated | | Observed | | Calculated | | Observed | |
	$K=24$ $\sigma=6 \times 10^9$	$K=25$ $\sigma=0$	Heavy rain	Watered by hose	$K=24$ $\sigma=6 \times 10^9$	$K=25$ $\sigma=0$	Heavy rain	Watered by hose
68	0·31	0·31	0·28	0·32	0·85	0·86	0·90	0.86
53·5	0·50	0·50	0·50	0·50	0·78	0·79	—	0·78
43·5	0·57	0·57	0·58	0·58	0·74	0·75	0·72	0·74

thus show that the site used was not abnormal. Smith-Rose and McPetrie (1934) have measured the constants of the soil at Teddington by a laboratory method at a frequency of 200 Mc/s. ($\lambda = 1.5$ m.), and they obtained values of $K=22$ and $\sigma = 3 \times 10^9$ e.s.u., and of $K = 4$ and negligible conductivity, under extreme wet and dry conditions respectively.

B. *Reflection and absorption from water*

Tests on the reflection of radiation of 9 cm. wave-length were made on fresh and salt water by placing a circular canvas dam, lent for the purpose by the National Fire Service, midway between the towers. This was filled to a depth of about 30 cm. with tap water or with a 4% solution of coarse salt, approximating to sea water. The diameter of the base of the dam was 5·5 metres (18 feet), and that of the water surface about 5 metres. The area available for reflection was consequently rather small, although experience had shown that it would probably be sufficient to give reasonably reliable results. The accuracy was, however, less than that of the ground-reflection experiments described above.

The absorption experiments were made using the same container as before, the cracks between the end plates and sides being caulked with Plasticine to render them watertight. The thickness of liquid under test was obtained by measuring the volume necessary to fill the container in each case.

(a) *Tap water.* The results obtained from the absorption tests on tap water are shown plotted in figure 3 (b), and indicate an attenuation of 380 db. per metre.

The most suitable values of K and σ to fit both reflection and absorption experiments were

$$K = 80, \qquad \sigma = 2 \times 10^{10} \text{ e.s.u.}$$

The reflection coefficients calculated from these values, together with the measured values are given in table 3.

(b) *4% salt solution.* The results from the absorption tests on a solution of 4% coarse salt in tap water are shown plotted in figure 3 (b), and indicate an attenuation of 1100 db. per metre.

Table 3.   Reflection coefficients for tap water, $\lambda = 9$ cm.

Angle of incidence (degrees)	Vertical polarization		Horizontal polarization	
	Calculated	Observed	Calculated	Observed
70	0·51	0·51	0·92	0·90
55	0·69	0·67	0·88	0·88
44·5	0·73	0·70	0·85	0·83

The most suitable values of K and σ to fit both reflection and absorption experiments were

$$K = 80, \qquad \sigma = 5 \cdot 5 \times 10^{10} \text{ e.s.u.}$$

The reflection coefficients calculated from these values, together with the measured values, are given in table 4.

Table 4.   Reflection coefficient for 4% salt solution, $\lambda = 9$ cm.

Angle of incidence (degrees)	Vertical polarization		Horizontal polarization	
	Calculated	Observed	Calculated	Observed
70	0·54	0·56	0·93	0·92
55	0·71	0·71	0·89	0·89
44·5	0·74	0·73	0·86	0·84

When making the absorption measurements on all types of material, observations were made using both radiation polarized parallel to and perpendicular to the edge of the container. It was found that a consistent difference of about 0·5 db. was present between the results with the two polarizations. A test on the empty container with the end plates removed showed that the container slightly increased the gain of the horn, and that this increase differed between the polarizations by the amount necessary to account for the discrepancy with the filled container. The experimental points plotted on the curves of figure 3 are all mean values of the results obtained with each polarization.

C. Reflection from turf-covered ground

A few measurements were made on the reflection coefficient of ground covered with short grass. The results obtained showed that, in general, specular reflection was not present. An exception was found when observations were made on a recently cut and rolled cricket wicket. The distance between the mirrors during the test was 21·5 metres and the height of the aerials 1·85 metres each, giving an angle of incidence for the reflected radiation of 80°. The results obtained are given in table 5, which also gives calculated values of reflection coefficient based on suitable assumed values of K and σ. Values obtained on an adjacent cricket wicket on which the grass had been allowed to grow to a height of about 2 cm. are also included in the table.

The value of K corresponding to the wet condition is lower than might be expected from the bare-ground experiments, since the site appeared to be well

soaked by the previous night's rain. No opportunity occurred to take measurements under conditions such that the ground was certainly saturated. Accordingly,

Table 5. Reflection coefficients of level grass-covered ground (Cricket
wicket), $\lambda = 9$ cm. Angle of incidence, 80°.

Conditions	Vertical polarization		Horizontal polarization		Assumed values	
	Calculated	Observed	Calculated	Observed	K	σ (e.s.u.)
Cut very short and rolled, dry. (No rain for at least 7 days)	0·47	0·46	0·79	0·79	3	5×10^9
Cut very short and rolled, wet. (Several hours' rain in previous night)	0·36	0·37	0·86	0·86	6	5×10^9
Grass about 2 cm. high, wet. (Same day as previous test)	0·36	0·51	0·86	0·83	6	5×10^9

although the results appear to indicate that the value of K for ground covered with turf is less than that of bare ground of the same humidity, this conclusion cannot be given with certainty.

D. *Summary of specular reflection results*

The values of K and σ appropriate to the various reflecting surfaces used in the investigation are brought together, for convenience, in table 6, which also gives the attenuation in the medium.

Table 6

Nature of medium	K	σ (e.s.u.)	Attenuation (db./metre)
Very dry sandy loam	2	3×10^8	36
Saturated sandy loam	24	6×10^9	220
Tap water	80	2×10^{10}	380
4% solution of coarse salt	80	$5 \cdot 5 \times 10^{10}$	1100
Dry turf	3	(5×10^8) (estimated)	(50) (calculated)
Wet turf	6	(10^9) (estimated)	(80) (calculated)

The values of attenuation and conductivity found for salt and tap water are somewhat lower than those previously obtained, using a different method, by Turner for a wave-length of 10 cm., but agree with his results in showing that the ratio between the conductivities of salt and tap waters is much less than at low frequencies. The value of 80 for the dielectric constant of water is not precise; a figure of 75 would fit the experimental results equally well.

§4. EXPERIMENTAL RESULTS: DIFFUSE REFLECTION

At centimetre wave-lengths, specular reflection is likely to be the exception rather than the rule, since the reflecting surfaces normally encountered are neither level nor free of vegetation. Experiments have been made to determine the reflection coefficient of level ground covered with vegetation, and of uneven ground.

A. *Effect of vegetation*

The results given in the last line of table 5 show the effect of grass about 2 cm. high on the reflection coefficient of level ground at an angle of incidence of 80°. The effect of the grass was to decrease the reflection coefficient for horizontally polarized waves, and to increase it for vertically polarized waves. The latter result was most unexpected, but the phenomenon was observed on several occasions (see also table 7). On longer grass the reflection coefficient for both polarizations was reduced; unmown grass in May, approximately 30 cm. high, gave coefficients of 0·19 with horizontally and 0·13 with vertically polarized radiation, at an angle of incidence of 80°.

Table 7. Reflection coefficients of vegetation-covered wet ground, $\lambda = 9$ cm.

Appearance of site, and height of vegetation		Angle of incidence of radiation, θ					
Appearance	Height (cm.)	$\theta = 68°$		$\theta = 53°·5$		$\theta = 43°·5$	
		V	H	V	H	V	H
Bare	0	0·32	0·86	0·50	0·78	0·58	0·74
True leaves beginning to form, ground visible	3–4	0·40	0·50	0·44	0·55	0·47	0·56
Dense clumps, ground showing in places	9–12	0·18	0·65	0·23	0·58	0·33	0·49
Ground almost obscured	20–25	0·06	0·32	0·10	0·39	0·17	0·41
Ground completely obscured	35–45	0·04	0·19	0·05	0·26	0·11	0·28

The main experimental results were taken on the site between the towers, on which the bulk of the previous results on ground reflection had been made. The levelled area was sown fairly thickly with mustard seed (very approximately 1000 germinations per square metre) to obtain a quick-growing carpet of vegetation. To ensure constancy of ground conditions, the site was watered thoroughly before each test as described in § 3, A (b). The results are given in table 7 for both vertical, *V*, and horizontal, *H*, polarizations.

The site at the conclusion of the test was covered completely with a dense carpet of vegetation, the top surface of which was comparatively level. It is probable that the reflection coefficient would be even lower with a less uniform ground covering. These results are in general agreement with those given by other investigators.

B. *Effect of uneven ground*

The effect of an uneven ground surface was investigated by raking the bare surface of the site into a series of uniform ridges. When planning the experiments,

the possibility of obtaining information applicable to the surface of the sea was borne in mind, and the system of ridges was made to simulate waves. Tests were made with these running along, across, and at 45° to the direction of transmission, using distances of 0·6 and 1·2 metres between successive crests, and various depths of between 5 and 16 cm. between crest and trough. Table 8 shows the reflection coefficients observed when the ridges were at 45° to the direction of transmission, while table 9 gives those obtained when the ridges were either along or across the direction of transmission. In these tables D gives the distance (cm.) between the crests of successive ridges, and h the mean depth (cm.) between crest and trough, while the column marked "Level" gives the estimated reflection coefficient for level ground of the same humidity as that under test.

Table 8. Reflection coefficients of ground ridged at 45° to the direction of transmission, $\lambda = 9$ cm.

Angle of incidence (degrees)	Vertical polarization				Horizontal polarization			
	Level	$D=60$ $h=14$	$D=120$ $h=10$	$D=120$ $h=5$	Level	$D=60$ $h=14$	$D=120$ $h=10$	$D=120$ $h=5$
68	0·08	0·07	0·09	0·13	0·65	0·14	0·18	0·30
53·5	0·13	0·04	0·05	0·07	0·51	0·04	0·06	0·16
43·5	0·22	0·04	0·04	·0·04	0·45	0·07	0·06	0·10

Table 9. Reflection coefficients of ground ridged along or across the direction of transmission, $\lambda = 9$ cm.

Angle of incidence (degrees)	Vertical polarization				Horizontal polarization			
	Level	Along $D=60$ $h=14$	Across $D=60$ $h=16$	$D=120$ $h=12$	Level	Along $D=60$ $h=14$	Across $D=60$ $h=16$	$D=120$ $h=12$
78	0·23	0·2	0·1	0·3	0·86	0·4	0·2	0·4
68	0·06	0·03	0·05	0·12	0·76	0·07	0·10	0·18
53·5	0·28	0·03	0·02	0·06	0·64	0·10	0·12	0·16
43·5	0·36	0·04	0·08	0·08	0·58	0·04	0·08	0·14

The results show clearly how small an irregularity in the ground surface is sufficient to prevent specular reflection; the reflection coefficient, moreover, becomes erratic with respect to angle of incidence when it has fallen below a value of about 0·1. The values given for level ground, although consistent among themselves, only relate very approximately to the state of the ground in the ridged condition, since the length of time necessary to make all the measurements, several days, was sufficient to allow marked changes in ground humidity. The difference between the values for level ground given in table 8 and those in table 9 are due to the observations recorded in the former table having been taken later in the year than those recorded in table 9, when the ground was appreciably drier. Values are given in table 9 for an angle of incidence of 78°. These were obtained with the mirrors at ground level about 14 metres apart. The accuracy

in these circumstances was not considered sufficient for measurements under conditions of specular reflection, but is adequate to give an indication of the reflection present from uneven ground.

The presence of a pattern in the signal reflected from ridged ground had been expected at the start of this investigation. Measurements were therefore made in which the mirrors, although always maintained at a distance of 15 metres apart, were moved so that the point of reflection to be expected from geometry was successively at different parts of the ridge system. Thus, when working on ridges running across the direction of transmission, the mirrors were moved together in the line of transmission to a number of successive positions so that the geometrical point of reflection varied from being on the crest of a ridge to being in the trough between two ridges. The somewhat unexpected result was obtained that the reflection coefficient remained substantially constant, at each angle of incidence, provided that the mirrors were set each time to obtain a signal maximum. The variable strength of reflected signal obtained from the constantly changing surface of the sea, therefore, may be partly due to the impossibility of maintaining a mirror, at every instant, in optimum adjustment. With ridges 120 cm. apart, running along the direction of transmission, it was found, however, that the reflected signal with the geometrical point of reflection on the crest of a ridge was appreciably greater than when the point of reflection lay in a trough.

When working on ridged ground, multiple reflections were, in general, observed. Thus, when the ridges were across the direction of transmission, in addition to a maximum of reflected signal obtained with the mirrors set at approximately the elevations indicated by geometry, further maxima could be obtained by adjusting the mirrors so that one had a greater and one a less angle of elevation than the geometrical. Three maxima of received signal of approximately equal value, but having different settings of mirror elevations, were not unusual. With the ridges running along the direction of transmission, a number of maxima of received signal could similarly be obtained by suitable adjustments of the azimuth settings of the mirrors.

The experiments on ridged ground described above were made on a surface of comparatively low reflection coefficient, and some similar experiments were therefore made on a surface of greater reflecting power. For this purpose a mat of 1·2 cm. ($\frac{1}{2}''$) mesh galvanized wire netting was laid between the towers, and its surface formed into a series of ridges. The material was somewhat intractable, but by pegging it to the ground at frequent intervals in the troughs, a system of ridges closely similar to those used for the ground experiments was obtained. The results obtained are given in table 10.

The reflection coefficients were notably greater than those for similarly ridged ground, but bore much the same ratio to the level values as held in the case of bare ground. Absorption tests showed that the wire netting transmitted 0·57 or 0·53 of the radiation incident normally on it, according as the electric vector was polarized parallel or perpendicular to the edge of the netting. This high proportion transmitted accounts for the comparatively low reflection coefficient with vertical polarization when the netting was level, and raises some doubts whether the result as a whole can be taken as typical for a surface of high reflecting power.

Table 10. Reflection coefficient of wire netting ridged at 45° to the direction of transmission, $\lambda = 9$ cm:

Angle of incidence (degrees)	Vertical polarization			Horizontal polarization		
	Level	$D=60$ cm. $h=10$ cm.	$D=120$ cm. $h=15$ cm.	Level	$D=60$ cm. $h=10$ cm.	$D=120$ cm. $h=15$ cm.
68	0·80	0·20	0·25	0·99	0·25	0·31
53·5		0·20			0·31	
43·5		0·23			0·26	

C. Other observations on rough ground

When the investigation started, in February 1944, the site between the towers consisted of rough unmown land, having variations in ground level of about ± 7 cm. on the mean over the area under test, and covered with a dense matted layer of grasses about 30 cm. high, with dead stalks rising to a height of more than 60 cm. The reflection coefficient of this surface was measured, and the grass then mown as short as the roughness of the ground permitted. The ground contour of the central 6 × 8 metres of the area between the towers was measured, and a

(a) Before removing turf. (b) After removing turf.
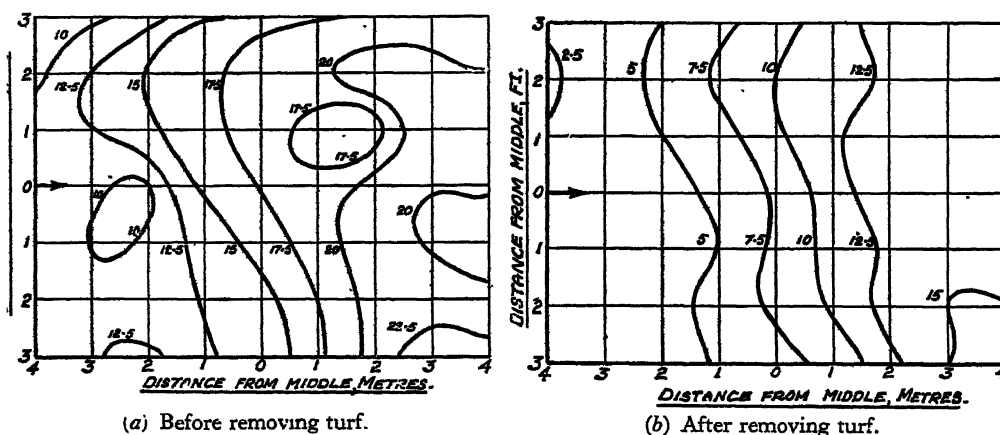
Figure 4. Ground contour maps of site.
Contours every 2·5 cm. above datum. Direction of transmission →

contour map is given in figure 4 (a). The reflection coefficient of this surface was also measured. The turf was next removed, and in doing this some of the ground irregularities were removed. A contour map of the site in this condition is given in figure 4 (b). Reflection coefficients were again measured on this surface.

The results of these tests are given in table 11, in which are also given the estimated reflection coefficients for level ground of the same humidity.

§5. DISCUSSION OF RESULTS AND CONCLUSIONS

The investigation has provided a fairly comprehensive outline of the phenomena associated with the reflection of radiation of 9 cm. wave-length, with two limitations. The first of these is the possibility of a scale effect when the results of

the experiments on diffuse reflection, made on an area of the order of 20 square metres, are applied to cases of transmission over considerable distances, when the area contributing to the reflected signal will be much larger.   The second limitation is that no results have been obtained with angles of incidence greater than 80°, since the method used is clearly not applicable to cases where the angular separation of the direct and reflected rays is small.

The general result of the experiments on diffuse reflection is to confirm the optical rule that, for regular reflection for a rough surface, the depth of the surface irregularities multiplied by the cosine of the angle of incidence must be a small

Table 11.   Reflection coefficients of rough ground, $\lambda = 9$ cm.

Angle of incidence (degrees)	Vertical polarization				Horizontal polarization			
	Level (estimated)	Long grass	Short grass	Bare	Level (estimated)	Long grass	Short grass	Bare
68	0·20	0·08	0·20	0·23	0·82	0·12	0·74	0·81
53·5	0·40	0·06	0·17	0·27	0·73	0·12	0·43	0·50
43·5	0·48	0·06	0·16	0·26	0·68	0·03	0·36	0·46

fraction of the wave-length.   The results of tables 8 and 9 indicate that when this fraction is about 1/5, the reflected ray may be reduced to about one half its proper value, and that when the fraction is 1/2 a reflection coefficient greater than 0·1 is improbable.   There is no reason to suppose that these relationships will be greatly altered by either an increase in the area of reflecting surface or in the angle of incidence.

The following conclusions on the reflection and absorption of radiation of 9 cm. wave-length can be drawn from the experiments described above :—

(1) Except at grazing incidence, specular reflection will be unusual, since it can only occur at very level surfaces which are free of vegetation.

(2) The reflection coefficient of bare ground when specular reflection occurs will usually lie between the values computed by assuming negligible conductivity and a dielectric constant of either 5 or of 15, according to the wetness of the surface layer.

(3) The change in phase on reflection is about 180° for vertical polarization at angles of incidence appreciably greater than Brewster's angle, and negligible at angles of incidence appreciably less than Brewster's angle.

(4) The change in phase on reflection is about 180° for horizontal polarization at all angles of incidence.

(5) For a very dry sand loam the attenuation is 36 db. per metre, the conductivity 3×10^8 e.s.u., and the dielectric constant 2.

(6) For the same ground when saturated with water the attenuation is 220 db. per metre, the conductivity 6×10^9 and the dielectric constant 24.

(7) For tap water the attenuation is 380 db. per metre, the conductivity 2×10^{10} and the dielectric constant between 75 and 80.

(8) For a 4% salt solution, approximating to sea water, the attenuation is 100 db. per metre, the conductivity 5.5×10^{10} and the dielectric constant between 75 and 80.

(9) Specular reflection ceases to occur when the surface becomes uneven (as would be predicted by optics); and when the product of the depth of surface irregularity and the cosine of the angle of incidence is $\lambda/2$ or more, the reflection coefficient is unlikely to exceed 0·1 for all polarizations and angles of incidence.

(10) A layer of vegetation on level ground prevents specular reflection, even if it is only a few centimetres high; in this case the reflected signal with vertical polarization may be increased at high angles of incidence.

(11) An appreciable thickness of vegetation reduces the reflection coefficient at all angles of incidence to a value considerable below that obtained by specular reflection from the ground surface; if the layer of vegetation is more than 30 cm. high, the reflection coefficient is unlikely to exceed 0·2 at angles of incidence less than 80°.

(12) The probable value of reflection coefficient for rough ground covered with vegetation is about 0·1, except in the neighbourhood of grazing incidence.

§6. ACKNOWLEDGMENTS

The authors wish to express their thanks to the National Fire Service for the loan of equipment, and to acknowledge the help and advice which they received from Dr. R. L. Smith-Rose in the planning of the investigation. The work was carried out as part of the programme of the Radio Research Board, and this paper is published by permission of the Department of Scientific and Industrial Research.

APPENDIX

Calculated reflection coefficients are given in table 12 for surfaces consisting of very dry ground, very wet soil, and sea water. The values for fresh water will only differ by about 0·03 from those for sea water.

Table 12. Calculated reflection coefficients, $\lambda = 9$ cm.

Angle of incidence (degrees)	Very dry soil		Very wet soil		Sea water	
	Vertical polarization	Horizontal polarization	Vertical polarization	Horizontal polarization	Vertical polarization	Horizontal polarization
0	+0·17	−0·17	+0·67	−0·67	+0·81	−0·81
20	+0·16	−0·19	+0·65	−0·68	+0·80	−0·82
40	+0·10	−0·24	+0·60	−0·73	+0·76	−0·85
60	−0·06	−0·38	+0·42	−0·82	+0·66	−0·90
70	−0·21	−0·51	+0·27	−0·87	+0·54	−0·93
80	−0·49	−0·71	−0·08	−0·93	+0·26	−0·97
85	−0·71	−0·84	−0·39	−0·96	−0·11	−0·98

In the table a + sign indicates that the phase change on reflection is very small, and a − sign indicates a phase change of about 180°.

REFERENCES

GRACE, A. C., 1946.   " A Note on a Radio Field-strength Measuring Set for use in the Frequency Range 400 to 4000 Mc./s."   *J. Inst. Elect. Engrs.* (Publication pending.)

McPETRIE, J. S., 1938.   " The Reflection Coefficient of the Earth's Surface for Radio Waves."   *J. Inst. Elect. Engrs.*, 82, 214.

SAXTON, J. A. and GRACE, A. C., 1946.   " A Field-strength Meter and Standard Radiator for Centimetre Wave-lengths."   (Publication pending.)

SMITH-ROSE, R. L., 1933.   " The Electrical Properties of Soil for Alternating Currents at Radio Frequencies."   *Proc. Roy. Soc.*, A, **140**, 359.

SMITH-ROSE, R. L. and McPETRIE, J. S., 1934.   " The Measurement of the Electrical Constants of Soil by a Lecher-wire Method at a Wave-length of 1·5 metres."   *Proc. Phys. Soc.*, **46**, 649.

NOTE ON THE BEHAVIOUR OF ZINC SULPHIDE PHOSPHORS UNDER CONDITIONS OF PERIODIC EXCITATION

BY MARY P. LORD * AND A. L. G. REES, †

Philips' Lamps, Ltd., Mitcham, Surrey

* Now at Imperial College, London.   † Now in Melbourne, Australia.

ABSTRACT.   The behaviour of luminescent solids under conditions of periodic excitation has been discussed theoretically and predictions made on the basis of simple models for the electronic processes accompanying the emission of luminescent radiation.   The significant features of the luminescence intensity as a function of time are the shift in phase with respect to the exciting radiation and the ratio of maximum to minimum emitted intensities.   The variation of these parameters with intensity and period of excitation can be used to distinguish between various mechanisms for the luminescence process.   Experiments are described, using a cathode-ray-oscillograph technique, illustrating the application of the method to zinc sulphide and zinc cadmium sulphide phosphors.   The general conclusions are that these phosphors do not possess any of the characteristics associated with the monomolecular mechanism, but show a semi-quantitative agreement with the requirements of a simple ionization-recombination mechanism.   Deviations from theory are ascribed to the complexity of these phosphors, which are activated by more than one type of activator atom.

§ 1. INTRODUCTION

MOST present-day applications of luminescent substances involve excitation by periodic stimuli, under which conditions phosphorescence is usually a desirable property; for example, the flicker in A.C.-operated fluorescent discharge lamps is reduced by the use of luminescent coatings exhibiting phosphorescence.   The behaviour of phosphors when subjected to periodic excitation has been studied very little, and no theoretical treatment of the effect has been attempted.   It has been observed, however, that the wave-form of the total light output from a fluorescent discharge lamp shows a slight advance in phase over the excitation wave-form (Davies, Ruff and Scott, 1942; Uyterhoeven and Zecher, 1938).   It was considered, therefore, that an experi-

mental study of the wave-form of the luminescent radiation alone might give some indication of the laws governing the effect and eventually provide some data concerning the mechanism of the luminescent process. This paper describes a limited experimental study of the effect, and compares the results with predictions based on a theoretical treatment. The results can be considered only as preliminary, but they do serve to indicate the possibilities of this method, particularly with regard to the mechanism of the luminescent process.

Wood (1921), Gottling (1923) and Szymanowski (1935) studied certain luminescent materials by a method similar in principle to that described here, but based on the use of the electro-optical Kerr effect for the measurement of very short time intervals. Both Wood and Gottling interpreted their results as indicating a dark interval between absorption of radiation and emission of luminescence, and it remained for Gaviola (1925) to show that this interval corresponded to the shift in phase of the luminescence intensity-time function with respect to that of the high-frequency spark excitation.

§2. THE BASIC LUMINESCENCE PROCESSES AND PERIODIC EXCITATION

It is to be expected that the behaviour of a phosphor under periodic excitation will depend markedly on the nature of the mechanism of the decay of luminescence. This mechanism is, in general, complicated (Randall and Wilkins (1945 a and b), Garlick and Wilkins (1945), Lord, Rees and Wise (1946)), but can be broadly classified into two main types: that in which electrons do not leave the luminescent centres, and that in which they do. In this preliminary investigation we have therefore examined the differential equations describing the luminescent emission of radiation under periodic excitation in the simplest case of each of the two types, and have noted certain significant differences in the solutions. The simplest case of the first type is that which for constant excitation leads to a simple exponential decay law. The simplest case of the second type is that which for constant excitation leads to a simple bimolecular decay law. More complicated cases of the two types are then discussed in a semi-quantitative way in § 3.

All periodic functions $I(t)$ representing the time-dependence of the intensity of the exciting radiation can, if necessary, be expressed as Fourier series, but it is evident that the simplest possible function, namely, that describing a sinusoidal variation, should be employed in the initial theoretical considerations, and also, if it is proved to be practicable, in any experimental work which attempts to establish the general predictions of the theory.

In the simplest case, in which the electronic transitions responsible for the absorption and emission of radiation in phosphors are all internal, i.e. in the case leading, for constant excitation, to a simple exponential decay law, the differential equation of the process is

$$\frac{dN}{dt} = \frac{\mu I(t)}{h\nu} - \alpha N,$$

where $I(t)$ is the energy of frequency ν incident per second, μ the absorption coefficient of the phosphor crystals, N the number of electrons in the excited level at any time t, and α the probability per unit time of a spontaneous transition

from the excited level to the ground state accompanied by the emission of radiation. For sinusoidal excitation, $I(t)$ has the form

$$I_m\left(1-\cos\frac{2\pi t}{T}\right),$$

where I_m is the maximum value of $I(t)$ occuring at $t=1/2T$, $3/2T$, etc., and T the period, so that the differential equation becomes

$$\frac{dN}{dt}=\frac{\mu I_m}{h\nu}\left(1-\cos\frac{2\pi t}{T}\right)-\alpha N.$$

This is a standard differential equation, the solution of which is

$$N=Ce^{-\alpha t}+\frac{\mu I_m}{\alpha h\nu}-\frac{\mu I_m}{\alpha h\nu}\left[\frac{\cos\dfrac{2\pi t}{T}+\dfrac{2\pi}{\alpha T}\sin\dfrac{2\pi t}{T}}{1+4\pi^2/\alpha^2T^2}\right],$$

C being the arbitrary constant of integration.

The "life" of the excited state, τ, the time taken for the number of atoms in such a state to decrease to $1/e$ of the original value in the absence of further excitation, is equal to $1/\alpha$, and, unless the excited state is accompanied by a metastable level, this can hardly be greater than 10^{-5} sec. This means that for large times $e^{-\alpha t}$ is a negligible term in the expression as α is 10^5 sec.$^{-1}$ or greater. The luminescence intensity is then described by the expression

$$\alpha N=\frac{\mu I_m}{h\nu}-\frac{\mu I_m}{h\nu}\left[\frac{\cos\dfrac{2\pi t}{T}+\dfrac{2\pi}{\alpha T}\sin\dfrac{2\pi t}{T}}{1+4\pi^2/\alpha^2T^2}\right],$$

This is a periodic function having a period equal to that of the exciting radiation. The phase shift of this function with respect to that describing the exciting radiation can be obtained by equating $\dfrac{d(\alpha N)}{dt}$ to zero and solving for

$$\epsilon=\frac{2\pi t_{\mathrm{max}}}{T}.$$

The solution is

$$\frac{2\pi}{T}\sin\epsilon-\frac{4\pi^2}{\alpha T^2}\cos\epsilon=0,$$

or $\qquad\qquad \tan\epsilon=\dfrac{2\pi}{\alpha T}.$

The significant feature of this result is that the phase shift ϵ is independent of the intensity of the exciting radiation. The ratio of intensities at maxima and minima of αN, as obtained by substitution of this value of ϵ in the original expression, is

$$\frac{1+\dfrac{\alpha T}{\sqrt{\alpha^2T^2+4\pi^2}}}{1-\dfrac{\alpha T}{\sqrt{\alpha^2T^2+4\pi^2}}}\qquad\text{or}\qquad \cot^2\tfrac{1}{2}\epsilon.$$

This also is independent of the intensity I_m, and is determined only by the period of the excitation and the "life" of the excited electronic state of the activator atom in the phosphor crystal. A plot of the phase shift against the ratio of maximum to minimum emitted intensities is given in figure 1.

In the simplest case of an ionization-recombination mechanism, i.e. in the case leading, for constant excitation, to a simple bimolecular decay law, the differential equation describing the process, under the periodic conditions of excitation as before, is

$$\frac{dN_c}{dt} = \frac{\mu I_m}{h\nu}\left(1-\cos\frac{2\pi t}{T}\right) - \beta N_c^2,$$

where N_c is the number of electrons in the conduction band at time t and β is the probability per unit time that an electron in the conduction band will encounter
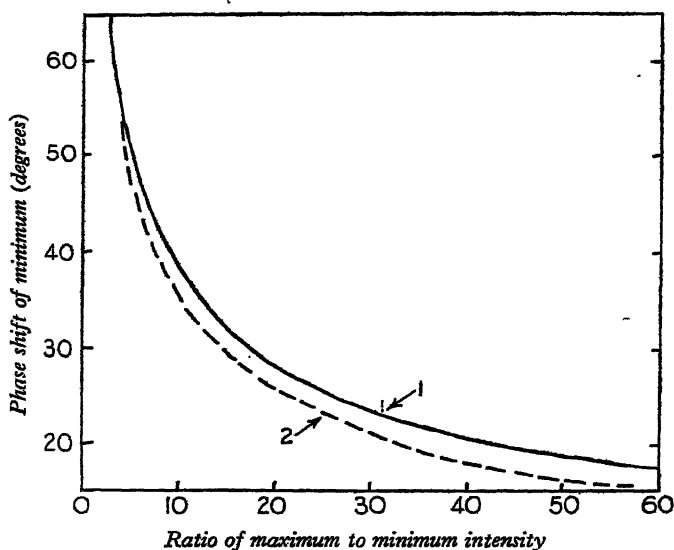
Figure 1. Theoretical relationships between phase shift of minimum luminescence intensity and ratio of maximum to minimum intensities under sinusoidal excitation.
1. Simple ionization-recombination mechanism. 2. Simple monomolecular mechanism.

a positive hole on an activator atom and recombine with it with the emission of luminescent radiation. The other symbols in the expression are defined as in the previous case.

The differential equation is not of a standard type; it is a particular form of the Riccati equation, which may be transformed into a linear form known as Hill's equation. A numerical solution has been given by Wise (in course of publication) in terms of the parameter

$$\kappa^2 = \frac{\beta\mu I_m T^2}{h\nu\pi^2},$$

and has been tabulated for values of κ^2 between 0·25 and 64. The solutions are given in the form

$$n^2 = \left(\frac{\beta N_c T}{\pi}\right)^2 = a + b'\sin\left(\frac{2\pi t}{T}+\phi'\right) + b''\sin\left(2.\frac{2\pi t}{T}+\phi''\right) + \ldots,$$

where a, b', b'',.... and ϕ', ϕ'',.... are determined by the value of κ^2. It is significant that the solution in this case depends on the maximum exciting intensity and hence, also, will the phase and the ratio of maximum to minimum emitted intensities. In table 1, the phase of the maxima and minima of βN^2 in terms of t, and the ratio of maximum to minimum luminescent intensity have been recorded for nine values of κ^2 between 0·25 and 64. These were obtained by a method of successive approximations from the numerical solutions given by Wise; the

Table 1

κ^2	Phase of minimum	Phase of maximum	Maximum luminescent intensity / Minimum luminescent intensity
→0	→90°	→270°	→1
$\frac{1}{4}$	65° 26′	242° 18′	2·47
$\frac{1}{2}$	59° 13′	233° 30′	3·40
1	50° 30′	221° 6′	4·85
2	43° 34′	211° 0′	6·87
4	38° 30′	201° 14′	10·0
8	32° 5′	194° 3′	14·2
16	26° 22′	189° 49′	21·3
32	21° 50′	187° 0′	35·4
64	17° 25′	186° 30′	63·0
→8	→0	→180°	→8

limiting values are also given. The theoretical relationships between the phase shifts at the maximum and minimum and the ratio of maximum and minimum values of βN_c^2 have been plotted in figure 1.

The form of the solution here is in marked contrast to that of the mono-molecular case, and the predictions embodied in these solutions can be used to obtain certain information simply and rapidly about phosphors whose electronic processes during absorption and emission of radiation are uncertain or unknown.

§3. MORE COMPLEX SYSTEMS: ELECTRON TRAPS

There are several ways in which the electronic processes leading to the emission of luminescence radiation may be complicated; the most important of these are (i) the presence of more than one type of activator in the phosphor, and (ii) the existence of metastable levels, associated with abnormalities in the crystal, which are capable of trapping electrons. The differential equations describing such systems are more complex: for the case of constant excitation intensity the theoretical treatment has been developed to a stage where the mechanism proposed can be said to describe the behaviour of certain phosphors in a qualitative way (Lord, Rees and Wise, 1946). It is evident that many phosphors considered to contain only one activator type did, in fact, contain more than one, and it is maintained that the failure to fit the simple exponential and bimolecular laws to experimental observations resulted, in part at least, from such a simplifying assumption.

Neglecting the presence of electron traps for the present, we may examine the case of a phosphor, containing more than one activator type, whose absorption

and emission processes are monomolecular in nature. We note that these processes at each activator atom will be independent and that the solution of an equation of the type

$$\frac{dN_r}{dt} = \frac{\mu_r}{h\nu} I(t) - \alpha_r N_r$$

is involved for each type of centre. The total luminous intensity is then

$$\sum_{r=1}^{n} \alpha_r N_r = \sum_{r=1}^{n} \frac{I_m \mu_r}{h\nu} \left\{ 1 - \frac{\alpha_r^2 T^2}{\alpha_r^2 T^2 + 4\pi^2} \left(\cos\frac{2\pi t}{T} + \frac{2\pi}{\alpha_r T} \sin\frac{2\pi t}{T} \right) \right\}.$$

The significant features of this function, namely, the phase and the ratio of maximum to minimum intensities, are, as in the case of a single type of centre, independent of the intensity of the exciting radiation.

For the ionization-recombination mechanism the solution of a set of simultaneous differential equations is required:

$$\frac{dN_r}{dt} = \frac{\mu_r}{h\nu} I(t) - \beta_r N_r N,$$

where N_r is the number of positive holes of the rth kind in the crystal, $N = \sum_{r=1}^{n} N_r$ the total number of electrons in the conduction band at time t, β_r the probability per unit time that an electron will recombine with a positive hole of the rth kind, μ_r the absorption coefficient for centres of the rth kind. The solution of this set of equations has not yet been attempted, but it is certain that the phase shift and ratio of maximum to minimum emitted intensities will depend on the intensity of the excitation.

A complete treatment of the problem should also take into account the variation with depth of the conditions in the phosphor (Lord, Rees and Wise, 1946), but the solutions of the appropriate equations have not yet been obtained. It is not, however, anticipated that the qualitative conclusions will be modified.

The existence of "electron traps" in many crystal phosphors has been established with certainty (Randall and Wilkins, 1945 a and b), and it would be instructive to examine the way in which these traps influence the luminescence processes. For the case of constant excitation intensity in electronic processes involving the conduction band, some progress has been made towards a solution of the problem (Lord, Rees and Wise, 1946), but for periodic excitation conditions it is somewhat more complicated. However, when the period T is short, 10^{-2} sec. or less, it is expected that electron traps will play a relatively small part in the luminescence processes, and that they will not contribute appreciably to the emission, except in those systems where the electron must pass through the metastable or trapping level before reaching the level from which a transition may be made to the ground state with the emission of radiation.

Systems in which the electron is excited from atoms of the crystal matrix, followed by a migration of the positive hole to the activator atom, where the electron-positive hole recombination takes place, have not been considered here, as the case of constant excitation intensity for such phosphors has not yet been treated theoretically.

It therefore appears that it should be possible to determine the general type of luminescence mechanism followed by a phosphor by examination of its behaviour under short-period excitation. The derivation of such information is illustrated in the subsequent sections of this paper.

§4. MEASUREMENT OF TIME DEPENDENCE OF LUMINESCENCE INTENSITY

The phosphors employed in these experiments were;

(i) ZnS, activated by 7.5×10^{-3} % Ag, but also containing interstitial Zn and a trace of Cu ($\leqslant 1 \times 10^{-4}$ %). The phosphor therefore contained three types of activating centre (see Lord, Rees and Wise, 1946).

(ii) ZnCdS containing 18% ZnS, activated by 1×10^{-2} % Cu, and presumably containing interstitial Zn and Cd also.

It was found that the intensity of the group of lines at $\lambda 3650$ from the quartz inner tube of an 80-watt MB/V high-pressure mercury-vapour lamp operated on 50 cycles varied in a substantially sinusoidal manner with time. Such a lamp constituted a convenient source of exciting radiation and was used for the experiments on the zinc and zinc-cadmium sulphide phosphors.

For recording the luminescence intensity-time relation, a cathode-ray-oscillograph method was employed. Light from the source, enclosed in a ventilated lamp-house, was passed through an 18A Wratten filter to isolate the $\lambda 3650$ radiation and then on to the phosphor, coated uniformly on a thin glass plate. Owing to the proximity of the lamp, the temperature of the phosphor was about $30°$ c. The transmitted radiation consisted of unabsorbed $\lambda 3650$ and emitted luminescence, and the former must be removed by an appropriate filter before the radiation strikes the photosensitive surface of the detector. An electron-multiplier photocell (Cinema Television, Ltd., MA 20, No. 300) was used as detector, the output from which was passed through a 1·2 megohm resistor and the p.d. appearing across this resistor applied directly to the Y-plates of a Mullard GM 3156 cathode-ray oscillograph. Synchronization is automatic if the time base and light source are both actuated by the same A.C. supply. A biasing voltage of $+48$ volts was required to bring the zero line to a convenient part of the oscillograph screen. The H.T. supply (*ca.* 1000 volts) for the multiplier was obtained from a rotary D.C. transformer operating on 14 v. D.C.; the generator output was smoothed by means of conventional filters before application to the multiplier potentiometer. Rigorous precautions were taken to exclude stray fields from the oscillograph.

The light filter system between the phosphor and the multiplier receiver must allow only that emitted light which it is proposed to investigate to fall on the photosensitive surface. For ZnS.Ag, whose emission is predominantly in the blue, a Wratten No. 39 (blue) filter, to remove green phosphorescence due to Cu, and a glass cell containing saturated aqueous sodium nitrate, to remove transmitted $\lambda 3650$, were employed; for ZnCdS.Cu, whose emission maximum lies in the orange-red, a Wratten No. 8 (yellow) filter sufficed to remove both transmitted $\lambda 3650$ and visible blue or green radiation.

Calibration of the vertical displacements in the oscillograph of the zero line and the trace with respect to exciting intensity proved necessary, since the zero

line was found to rise owing to a build-up effect in the 1·2 MΩ resistor, and the responses of the multiplier and oscillograph were not linear. Calibrations were performed with the mercury lines lying in the range of the spectral emission of the phosphor under investigation by reducing the intensity in known ratios with a calibrated neutral wedge.

The wave-forms of the luminescence for four different excitation intensities were photographed on the same plate together with the wave form of the Hg λ3650 radiation, used to fix the origin of $I(t)$. For each plate, the deflection caused by light reaching the photocell with the luminescent layer removed, but with filters in place, was photographed; however, this deflection proved to be negligible. The true zero line appropriate to each observed wave form was derived, the height above it of the maxima and minima obtained and corrected for non-linear response by means of calibration curves. All traces were recorded independently on three occasions, the traces measured from optical enlargements and the mean values taken. A typical photographic record is shown in figure 2. The absolute intensities of excitation radiation were obtained from measurements with a Moll thermopile.
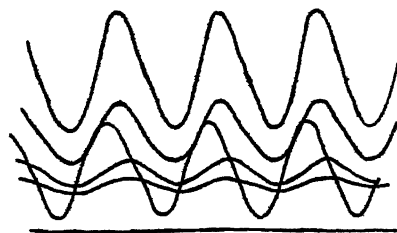
Figure 2. Reproduction of a typical oscillograph record showing phase shift.

§5. RESULTS AND DISCUSSION

The results of the experiments conducted are recorded in table 2 and plotted together with the theoretical curve for the region of the experimental observations, in figure 2.

Table 2

Phosphor	I (watts/cm²)	Ratio of maximum to minimum intensity (mean)	Phase shift of minimum (degrees) (mean)
ZnS . Ag	$7·0 \times 10^{-4}$	2·93	40·3
	3·5	2·18	43·7
	2·0	1·89	48·9
	1·0	1·53	51·7
ZnCdS . Cu	$7·0 \times 10^{-4}$	9·8	22·3
	3·5	8·3	24·4
	2·0	5·8	27·8
	1·0	4·3	30·7

Owing to the presence of a certain amount of distortion of the peaks near the top of the oscillograph screen, the phases were measured only at the minima.

From the curves in figure 3 it is apparent that the qualitative behaviour of the phosphors is as predicted by the theory outlined in §3, but the quantitative agreement is poor.   This is not suprising in view of the thickness of the layer, the complex nature of the systems constituting these phosphors and the possible effect of electrons trap at the relatively low frequency of excitation used (10^2 cycles per second).   However, the results demonstrate quite conclusively that the luminescence mechanism is not monomolecular in type since all points in figure 3 for one phosphor corresponding to different intensities would then be identical, and, indeed, lend support to the belief that it is, *ca.* 30° c., of an ionization-recombination type.   It is not necessarily expected that the deviations from theory for the two phosphors investigated should be the same; however, it can be seen from figure 3 that both sets of points approximate to the same curve. As the
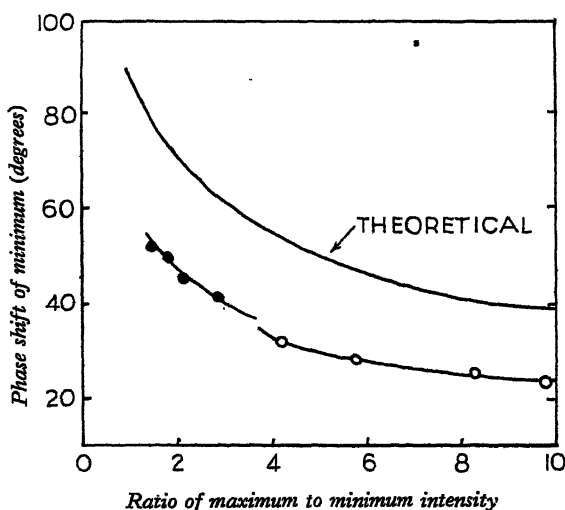
Figure 3.   Comparison of theory and experiment.   Theoretical curve from table 2 for ionization-recombination model.   Experimental observations: $\circ$ ZnCdS . Cu, $\bullet$ ZnS . Ag.

deviations are in all probability largely due to interstitial lattice atoms (Zn or Cd), whose equilibrium concentrations at the temperature of preparation of the phosphor would be of the same order, this fact is not surprising.   It is evident that all phosphors following the simple bimolecular mechanism should satisfy the theoretical relation of figure 1 irrespective of the value of the recombination coefficient, absorption coefficient, intensity or periodicity of the excitation.

It would appear that the principles of the method here discussed should be of considerable value in distinguishing between various models of the luminescence mechanism, since the interpretation of decay curves under constant excitation · conditions is a matter of some difficulty and uncertainty.

REFERENCES

DAVIES, RUFF and SCOTT, 1942. *J. Instn. Elect. Engrs.*, **89** (II), 447.
GARLICK and WILKINS, 1945. *Proc. Roy. Soc.*, A, **184**, 408.
GAVIOLA, 1925. *Z. Phys.*, **35**, 748.
GOTTLING, 1923. *Phys. Rev.*, **22**, 566.
LORD, REES and WISE, 1946. *Proc. Phys. Soc.*, **58** (in course of publication).
RANDALL and WILKINS, 1945 a. *Proc. Roy. Soc.*, A, **184**, 365.
RANDALL and WILKINS, 1945 b. *Proc. Roy. Soc.*, A, **184**, 390.
SZYMANOWSKI, 1935. *Z. Phys.*, **95**, 440.
UYTERHOEVEN and ZECHER, 1938. *Philips' Tech. Rev.*, **3**, 272.
WISE. *Phil. Mag.* (in course of publication).
WOOD, 1921. *Proc. Roy. Soc.*, A, **99**, 362.

NOTE ON THE RAPID DETERMINATION OF DECAY CHARACTERISTICS OF LUMINESCENT SOLIDS

BY MARY P. LORD* AND A. L. G. REES,†

Philips' Lamps, Ltd., Mitcham, Surrey

* Now at Imperial College, London.   † Now in Melbourne, Australia.

ABSTRACT. For the provision of rapid and convenient means for the examination of luminescent solids by oscillographic methods, the possibility of producing square-wave light pulses from a gaseous discharge has been investigated. Decay curves sufficiently accurate for the estimation of half-life periods and for the study of gross effects could be obtained by interrupting the D.C. supply to the discharge by means of a small vibrator.

FOR comparative and routine purposes, a sufficient criterion of the decay of luminescence of a phosphor is its " half-life ", that is, the time taken for the emitted intensity to fall to half its value after the removal of excitation. Methods less tedious and precise than the orthodox have therefore been developed to derive decay curves rapidly.

An electron-multiplier phototube together with a cathode-ray oscillograph was first used for the plotting of decay curves by de Groot (1939). His method involved the use of rotating sectors for the production of square pulses of exciting radiation; the apparatus described here provides for the production of these square pulses directly from the gaseous discharge. The specific advantage of this method lies in the absence of motor-driven moving parts, leading to a more convenient and portable equipment; it also avoids the provision of an adjustable constant-speed motor and provides simpler synchronization for the oscillograph time-base. Use was made of the same multiplier-oscillograph combination as that employed in the study of the behaviour of sulphide phosphors under periodic excitation (Lord and Rees, 1946). We were concerned here, however, with the examination of silicate materials and, therefore, used a low-pressure mercury discharge as the source of λ2537 radiation.

Square wave-forms were produced by the following methods: (i) The use of a bright-emitter diode run under saturation conditions; (ii) the use of a thyratron in series with a saturated diode; (iii) interruption of the D.C. supply to the discharge with a mechanical vibrator. The first method did not give a completely satisfactory flat-topped wave and could not be used as such; when, however, a
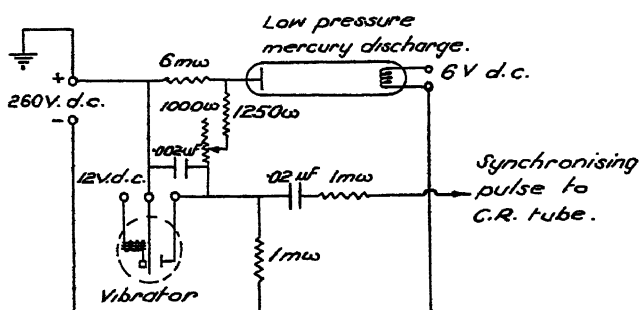
Figure 1.   Circuit for production of square-wave pulses for λ 2537 radiation.

thyratron was connected in series, the wave-form was improved and variation of the pulse duration by control of the grid bias then became possible.

As the current requirements of the discharge lamp were of the order of 70 ma. for satisfactory light output, the use of a mechanical vibrator (V 6121, 12 volt) was found to be simpler. This vibrator was used in the circuit outlined in figure 1 to interrupt the 260-volt D.C. mains, which were smoothed by means of a conven-
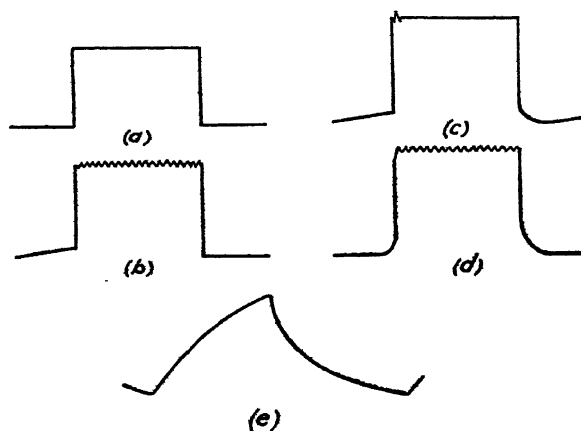
Figure 2.   Wave-form characteristics of vibrator circuit.

(a) Discharge current ; (b) lamp voltage ; (c) vibrator voltage ; (d) light output from discharge ; (e) build-up and decay of luminescence in manganese-activated zinc silicate.

tional pi filter. A synchronizing pulse was taken to the oscillograph through a condenser and resistor as shown. The 6-megohm series resistor ensured that a constant small number of ions were available to facilitate starting up the lamp on each cycle. For currents up to 70 ma. (the lamp rating), satisfactory wave-forms were obtained, but with some build-up and decay and a small-amplitude high-frequency oscillation superimposed on the flat top. The high frequency has its origin in the discharge, as can be seen from the wave-forms in figure 2, (a, b and c).

The current wave-form of figure 2 (*a*) was obtained by placing a series resistance of 150 ohms in the negative line and applying the voltage established across this resistor to the vertical deflector plates. Figure 2 (*b*) was obtained by applying the voltage across the lamp directly to the deflector plates, using a shunt of 250,000 ohms across the lamp. The vibrator voltage wave-form, shown in figure 2 (*c*), was obtained by connecting the vertical deflector plates across 50,000 ohms of the 250,000-ohm shunt; however, the discharge was suppressed in this case by removing the filament heating supply. Bounce of the contacts on make is evident at the beginning of the trace.

Curves showing the build-up and decay of luminescence, as in figure 2 (*e*), are obtained by interposing the phosphor layer and appropriate filter system between the square-wave source and the photomultiplier. Half-life periods deduced from such curves are in substantial agreement with those obtained by more precise methods (Lord, Rees and Wise, 1946), provided corrections are made from calibration for the non-linearity of the multiplier-cathode-ray oscillograph system.

The method may be extended to a wide range of pulse durations and repetition rates by substituting vibrators of different frequencies and times of contact. The vibrators do not perform satisfactorily if maximum current is drawn for long periods; the wave form deteriorates if the unit becomes heated.

Although neither this method nor that of de Groot can be used to obtain decay curves sufficiently precise to give useful data about the electronic processes accompanying luminescence, this method is quite adequate for the study of gross effects and for the rapid comparison of phosphors.

REFERENCES

DE GROOT, 1939. *Physica*, 6, 275.
LORD and REES, 1946. *Phys. Proc. Soc.*, 58, 280.
LORD, REES and WISE, 1946. *Proc. Phys. Soc.*, 58 (in course of publication).

BAND SYSTEMS IN THE SPECTRUM OF NITROGEN

By A. G. GAYDON, Imperial College, London

AND

(Mme) RENÉE HERMAN, Observatoire de Lyon, France

ABSTRACT. Several new band systems of N_2 which have been reported by us indepen-
dently during the war are compared. Several systems have been obtained by both of us,
and agreement is generally satisfactory. There appear to be eight progressions of bands
in the ultra-violet which are due to transitions to the $a\,^1\Pi$ state. Differing vibrational
analyses for the Fifth Positive (Van der Ziel) system are discussed, and it is shown that
Gaydon's interpretation is to be preferred, while for Kaplan's Second system (Herman's a)
Herman's vibrational analysis is more satisfactory. Bands in the green region are compared.

§1. INTRODUCTION

DURING the war we have independently investigated band spectra obtained
in emission from nitrogen in discharge tubes and have each found several
new systems (Gaydon, 1944 a, 1944 b; Gaydon and Worley, 1944; Herman,
1945). On the whole our results are in fairly good agreement, but there are some
ways in which our interpretations of the observations differ. The names which
we have applied to the various new systems are naturally quite different. Thus
a joint publication seems desirable in order to assist others to sort out the
numerous systems attributed to N_2 which have now been found and, as far as
possible, to clear up some of the differing interpretations of the observations.

Generally speaking, Herman has developed rather more intense sources, and
has thus been able to obtain some of the band systems more completely, so enabling
the more extensive and reliable vibrational analyses. Gaydon, on the other hand,
has used spectrographs of larger dispersion and has concentrated more on the
rotational structure of the bands; data obtained from rotational analyses of a
number of bands have led to definite conclusions on the nature of the electronic
levels.

The characteristics of the discharge-tube sources used for exciting the spectra
have been rather different, and it is not surprising to find that while some of the
new band systems have been observed by both of us, other systems have been
reported by only one of us.

§2. BAND SYSTEMS INVOLVING TRANSITIONS TO THE $a\,^1\Pi$ STATE

Many of the new systems consist of single progressions of bands showing the
same vibrational intervals as those of the $a\,^1\Pi$ state, which is the upper level of the
main Lyman system of N_2. This level was formerly believed to be $^1\Pi_u$ but
Herzberg (1946) has recently suggested that it may be $^1\Pi_g$. These band systems

thus consist of transitions from a number of highly excited levels to the $a\,^1\Pi$ level, this being confirmed in several cases by the rotational analyses. In table 1 we have set out brief particulars of these systems, with the names which we have given to them in out separate papers.

Table 1. Band systems with $a\,^1\Pi$ as final level

ν_0	(0,0)	Approx. λ (0,1)	(0,2)	Gaydon's designation	Herman's designation	Type of upper electronic state
35371	2827	2967	3119	P	ζ	$^1\Sigma$
35761*	2795	2932	3079	—	$\delta, v'=0$	
36394	2746	2878	3020	Q	$\delta, v'=1$	$^1\Pi$
37417	2672	2796	—	R	—	$^1\Sigma$
41705	2397	2497	2603	S	—	$^1\Sigma$
42373*	2359	2455	2558	—	θ	
43818	2281	2371	—	T	ϵ	$^1\Sigma$
46611*	—	2224	2308	—	η	

* These values are derived from head measurements ; other values of ν_0 are for band origins.

If the $a\,^1\Pi$ state is $^1\Pi_u$, then the $^1\Sigma$ states will be $^1\Sigma_g^-$, as given in Gaydon's paper, and the upper $^1\Pi$ state will be $^1\Pi_g$. If, however, the $a\,^1\Pi$ state is $^1\Pi_g$, then the $^1\Sigma$ states will all be $^1\Sigma^+$ and the upper $^1\Pi$ will be $^1\Pi_u$.

For the system with $\nu_0 = 35371$ our analyses agree; these bands appear more strongly on Gaydon's plates.

The two progressions with ν_0 35761 and 36394 were tentatively grouped together as one system by Herman. However, since the first progression is not present on Gaydon's plates, it now seems more probable that the two progressions belong to separate systems. The value of $\omega_{\frac12}$, about 640 cm.$^{-1}$, obtained by combining the two progressions, also seems rather small for a state with a B value of about 1·36. Herman obtained four bands in each progression. Gaydon obtained only the three strongest bands of the ν_0 36394 progression, but achieved rotational analyses of two of these. Janin (1943) has independently observed bands of both of these progressions, his measurements agreeing with ours.

The 37417 progression was only reported by Gaydon, who made a rotational analysis of the (0,0) band and also found the (0,1) band, which, it now appears, very nearly coincides with the (0,0) band of the 35761 progression. The (0,0) band at $\lambda 2672$ is, however, present on Herman's plates.

The $\nu_0 = 41705$ progression again was only reported by Gaydon, but some of these bands are also present on Herman's plates.

The 42373 progression was reported only by Herman, but there is some weak structure in the region of these bands on Gaydon's plates as well.

The 43818·progression·has been obtained by both of us. Gaydon discussed the possibility that this might be the $v' = 1$ progression of the system with $\nu_0 = 41705$, but decided the evidence was not sufficiently conclusive.

The progression with ν_0 at 46611 was reported only by Herman, but the not (0,1) and (0,2) bands at $\lambda 2224$ and 2308 are present on Gaydon's plates, although higher members of the progression are, absent or masked.

Thus there is evidence for the existence of eight independent systems of N_2 in the ultra-violet all having the common final level $a\,^1\Pi$. Most of these systems have been observed by both of us, although our original publications only gave three systems in common.

The close numerical coincidence between the energies of the upper levels of Herman's ϵ and ζ systems with those of the levels h and c obtained from absorption measurements by Watson and Koontz and by Birge and Hopfield led her to think that these levels, h and c, were the upper levels of the ϵ and ζ systems, thus indicating a violation of the selection rule that the symmetry u or g must change. Herzberg (1946) has now suggested that the $a\,^1\Pi$ state is of g symmetry and that the forbidden transition is the Lyman system itself, $a\,^1\Pi_g \rightleftharpoons x\,^1\Sigma_g^+$. This accounts for the great path length of gas required to produce the Lyman bands in absorption. It brings the upper level of the 43818 progression into coincidence with the level h, as suggested by Herman. The upper level of the 36394 progression would then also fit with Worley's level m, but the upper level of the 35371 progression could only be identified with the level c by assuming that the electronic type of this state c is $^1\Sigma_u^+$ and not $^1\Pi_u$, as formerly supposed; the B values agree satisfactorily. There is, therefore, much to be said for this suggestion by Herzberg, but it should be pointed out that it does not explain why strong absorption bands to the upper levels of our other progressions have not been reported.

§3. THE FIFTH POSITIVE OR VAN DER ZIEL'S SYSTEM AND KAPLAN'S SYSTEMS

Three other band systems of N_2 with related energy levels have also been studied by both of us. Our designations for these systems are given in table 2.

Van der Ziel originally reported two progressions of bands separated by an interval of 1827 cm^{-1}, and he assumed that these were the $v' = 0$ and 1 progressions of a system. The intensity distribution in this system on this assumption was obviously unsatisfactory. Gaydon obtained these two progressions on his plates and also found an additional progression separated by 1869 cm^{-1} from the

Table 2

Gaydon's designation	Herman's designation
Fifth Positive (van der Ziel)	Van der Ziel
Kaplan's First System	β
Kaplan's Second System	α

first of the older progressions. By treating the new progression as the $v' = 0$ and the original progressions as $v' = 1$ and 2, a satisfactory intensity distribution, in conformity with the Franck-Condon principle, was obtained. Rotational analysis of one band of the new progression proved that it belonged to the same system as the older progressions.

Van der Ziel's rotational analysis had shown that the bands were due to a transition between two Σ states, and Herman made an attempt to identify the upper state with that of a system of bands observed by Collins and Price in absorption, it being assumed that van der Ziel's two progressions arose from levels with

$v = 1$ and 3 of this state. By assuming preferential excitation to these levels a reasonable intensity distribution could be obtained. However, it is now found that Gaydon's additional progression is also present on Herman's plates, and as this cannot be fitted into Herman's scheme it seems that Gaydon's interpretation is to be preferred. Moreover, Gaydon points out that Herman's scheme would be inconsistent with the strong degradation of the bands to shorter wave-lengths, and that it is now known that the bands are due to a transition between two $^1\Sigma$ levels, whereas the upper level of Collins and Price's system is believed to be a triplet.

Kaplan originally reported a system with (0,0) band at 2153 A. and gave measures of six bands to the nearest angstrom, and suggested that they might involve a transition to the same final level as that involved in van der Ziel's system. Gaydon observed only three bands clearly, with a suggestion of a fourth, but was able to make an analysis of the (0,2) band, showing that the transition was $^1\Pi \rightarrow {}^1\Sigma$ and obtaining the value of B', and confirming that B'' had the same value as for the Fifth Positive system. Herman obtained a large number of bands (seven in all) and was able to show that $\omega_1' = 1708 \, \text{cm}^{-1}$.

Kaplan's second system of bands, which he also believed to have the same final level as van der Ziel's, has been observed by us. Gaydon only obtained the bands weakly and was unable to make a rotational analysis, and said that the vibrational intervals agreed as well as could be expected with those of the lower state of the Fifth Positive system. Herman, however, obtained this system more strongly, and recorded eleven bands, including new bands at $\lambda\lambda$ 2354 and 2263, which are also subsequently found to be present quite clearly on Gaydon's plates and which prove that the transition does not involve the lower state of van der Ziel's system, as had been thought by Kaplan and Gaydon. The vibrational energy interval for the upper state is about $1709 \, \text{cm}^{-1}$, which suggests that this system has the same initial $^1\Pi$ level as Kaplan's first system.

§4. OTHER SYSTEMS

For two other systems our respective designations are given in table 3.

Herman's γ system consists of two fairly long progressions, and the analysis shows that the transition is to the $A\,{}^3\Sigma$ level of N_2. Bands of this system are not present on Gaydon's plates. Kaplan had previously observed a number of these

Table 3

Gaydon's designation	Herman's designation
(Kaplan's third system)	γ
Green system	χ

bands and had suggested that some of them formed a progression involving the lower level of van der Ziel's system. This is now shown to be incorrect.

A system of rather complex bands in the green region was observed by Gaydon, who arranged seven bands into a tentative vibrational scheme. Herman also reported three bands at $\lambda\lambda$ 5594, 5330 and 5092 under small dispersion and referred to them as the χ system. More recently Herman has obtained this system with greater intensity, and revised measurements of five bands show clearly that this χ system is identical with Gaydon's Green system.

Mutual comparison of our results has led to the identification of a number of the unassigned bands on our plates, but there are still a number of weak bands, probably due to N_2, especially at the far end of the quartz ultra-violet region, which have not been analysed.

REFERENCES

GAYDON, A. G., 1944 a. *Proc. Roy. Soc.*, A, 182, 286; 1944 b. *Proc. Phys. Soc.*, 56, 85.
GAYDON, A. G. and WORLEY, R. E., 1944. *Nature, Lond.*, 153, 747.
HERMAN, R., 1945. Thèse (Paris : Masson).
HERZBERG, G., 1946. *Phys Rev.* (in the press).
JANIN, J., 1943. *Cahiers de Physique*, 16, 16.

DISCUSSION

Dr. S. P. SINHA. I would like to know the experimental conditions under which the new systems of nitrogen bands discovered by Dr. A. G. Gaydon were developed and if it was possible to explain their appearance in the light of the conditions maintained during the experiment.

AUTHOR'S reply. The experimental conditions for exciting the new band systems were fully described in earlier papers (Gaydon, 1944 a, 1944 b) and the conditions of excitation were discussed. Briefly, the ultra-violet systems were obtained by a condensed induction coil discharge through N_2 in a simple Pyrex tube, while the Green system required a Tesla discharge at relatively high gas pressure.

A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: VI. THE CHROMATIC VARIATION OF THE TANGENTIAL ABERRATIONS

By F. D. CRUICKSHANK,

University of Tasmania

ABSTRACT. Using those differential transfer coefficients of a lens system which specify the rate of change of the tangential aberrations with the refractive indices of the component singlet lenses of the system, a simple method is available for the rapid calculation of the chromatic variation of these aberrations. In a similar manner, other transfer coefficients (the chromatic coefficients of the system) provide a rapid means of obtaining a complete analysis of the secondary spectrum of the system at any zone in the case of the axial aberration, and at any obliquity of the transverse aberration. Examples of an actual system are given.

§1. INTRODUCTION

THE aberrations of an optical system are determined normally for incident light of some mean wave-length in the spectral range over which the system is expected to give satisfactory images. The full significance of the chromatic aberration of the system is really apparent only when, in addition to the usual analysis, the variation of the aberrations with the wave-length of the incident light

has been examined. Hitherto the analysis of the dependence of the aberrations on the wave-length of the incident light has required repeated ray-tracings in different colours. It is of interest, therefore, to examine this matter in the light of the properties of the transfer coefficients which have been developed in the earlier papers of this series.

Suppose that the aberrations of the system for incident light of some mean wave-length, d, have been obtained by a ray-trace and the general transfer coefficients have been calculated from the trace. If the mean refractive indices of the glasses of the components $a, b, \ldots h, \ldots$ of the system are changed by amounts $\delta N_a, \delta N_b, \ldots \delta N_h, \ldots$, the change in the value of the aberration A'_d which may be any of the tangential aberrations of the system, resulting from the glass changes is given by

$$\delta A'_d = \Sigma \frac{\partial A'_d}{\partial N_h} \delta N_h.$$

If, however, instead of considering glass changes we consider a change in the wave-length of the light which is incident on the system from the object, then the value of the aberration, A', for light of wave-length, r, will be given by

$$A'_r = A'_d + \Sigma \frac{\partial A'_d}{\partial N_h} (N_r - N_d)_h.$$

Hence, using the computed values of the transfer coefficients, $\partial A'/\partial N_h$, for the various aberrations, we can investigate the chromatic variations of the aberrations quite quickly.

§2. A TYPICAL EXAMPLE

As an example we consider a rough design of a wide-angle photographic objective of the Ross Xpres type. In table 1 the computed values of the transfer coefficients of the aberrations with respect to the refractive indices of the components are shown, the oblique aberrations being calculated from a pencil of about 22° obliquity.

Table 1

	LA'_T	LA_z	X'_T	$Coma'_T$	Dist'
$\partial/\partial N_a$	150·13	52·302	−100·48	0·891	7·780
$\partial/\partial N_b$	−140·51	−49·202	115·59	0·167	−7·029
$\partial/\partial N_c$	7·11	1·960	− 23·54	−1·887	1·382
$\partial/\partial N_d$	12·53	5·220	− 8·48	1·195	−1·322
$\partial/\partial N_e$	−115·64	−42·750	47·16	−0·651	5·529
$\partial/\partial N_f$	117·16	43·000	− 35·34	−0·240	−6·180

The glass selected for components a and f was DBC 615553, for components b and e LF 549467, and for components c and d HC 519604. The dispersions of these glasses are given in table 2. The ray-trace of the system having been carried out for sodium light, we calculate next the aberration differences, $A' - A'_D$, for wave-lengths from 7065 A. to 4341 A. Since the glasses in the front and rear halves of the lens are similar, we may combine the coefficients for the components which have similar glasses, as, in the evaluation of the differences by the equation given above, these coefficients will each be multiplied by the same dispersion values. In table 3 the full calculation is set out for the case of the marginal spherical

aberration, and finally, in table 4 the chromatic differences for all the aberrations are collected, the detailed computations being omitted.

The accompanying graphs in figures 1 to 3 show the variation of the aberrations plotted against the wave-length, and thus completely summarize the analysis of the
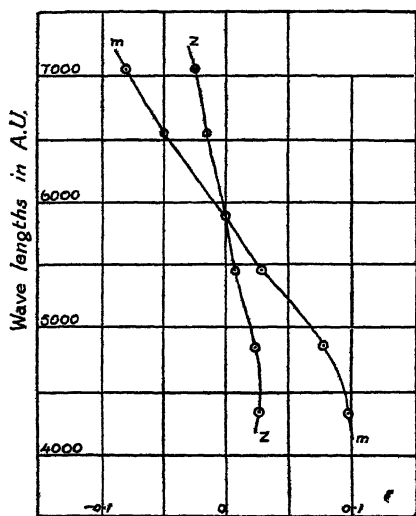
$LA' - LA'_D$

Figure 1.
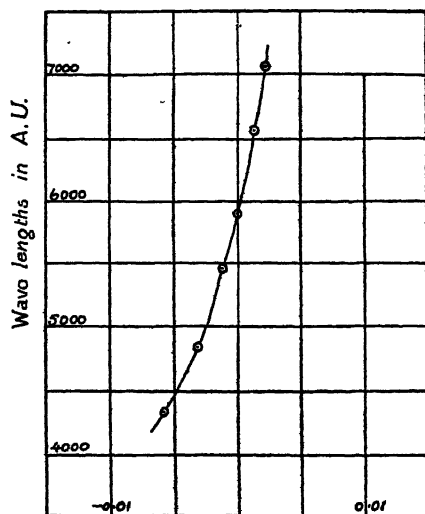
$Coma'_T - Coma'_{TD}$ (22 ° pencil)

Figure 2.
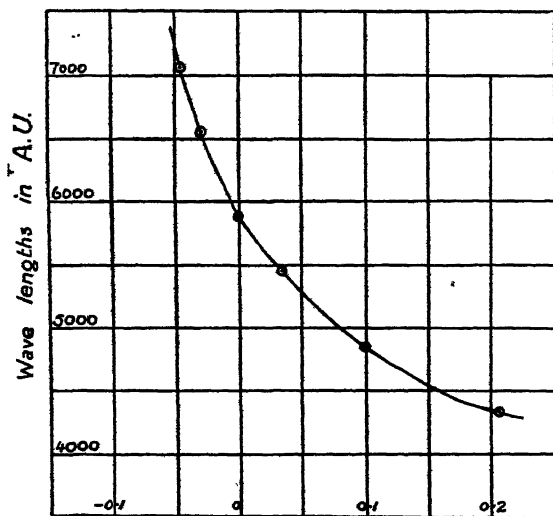
$X'_T - X'_{TD}$ (22 ° pencil)

Figure 3.

dependence of the general correction state of the system on the wave-length of the incident light. As a check on the general validity of this analysis, the differences of the spherical aberration for C and D light, and for F and D light, were determined by complete ray traces and compared with the values computed by the present

method. The results are given in table 5, the discrepancies being little greater than the uncertainty of the last figure of the traces.

Table 2

	DBC	LF	HC
N_b-N_D	−0·00517	−0·00539	−0·00406
N_C-N_D	−0·00328	−0·00342	−0·00256
N_e-N_D	0·00274	0·00291	0·00211
N_F-N_D	0·00785	0·00835	0·00603
N_G-N_D	0·01422	0·01529	0·01089

Table 3

Comp.	Coeff.	Contrbn. b–D	Contrbn. C–D	Contrbn. e–D	Contrbn. F–D	Contrbn. G′–D
a, f	267·29	−1·3819	−0·8767	0·7324	2·0982	3·8008
b, e	−256·15	1·3806	0·8760	−0·7454	−2·1388	−3·9164
c, d	19·64	−0·0798	−0·0503	0·0415	0·1185	0·2140
$LA'-LA'_D =$		−0·0811	−0·0510	0·0285	0·0779	0·0984

Table 4

Range	LA'_m	LA'_z	X'_T	$Coma'_T$	Dist′
b–D	−0·0811	−0·0263	−0·0450	0·00205	−0·00043
C–D	−0·0510	−0·0165	−0·0291	0·00129	−0·00027
e–D	0·0285	0·0087	0·0339	−0·00109	0·00015
F–D	0·0779	0·0236	0·0996	−0·00310	0·00040
G′–D	0·0984	0·0275	0·2082	−0·00568	0·00047

Table 5

	Marginal		Zonal	
	Trace	Coeffs.	Trace	Coeffs.
$LA'_C-LA'_D$	−0·066	−0·051	−0·018	−0·0165
$LA'_F-LA'_D$	0·071	0·0779	0·022	0·0236

§3. THE ANALYSIS OF THE SECONDARY SPECTRUM

To complete the discussion of the chromatic variation of the aberrations, the matter of the secondary spectrum should be considered. It has been shown previously that the final intersection length, L'_r, for an axial ray of some colour r is related to the corresponding quantity, L'_d, for the traced ray for some mean wave-length, d, by

$$L'_r = L'_d + \Sigma \frac{\partial L'_d}{\partial N_h} (N_r - N_d)_h,$$

and for a principal ray of an oblique pencil

$$H'_r = H_d + \Sigma \frac{\partial H'_d}{\partial N_h} (N_r - N_d)_h,$$

where H' measures the height at which the principal ray intersects a fixed plane, conveniently the paraxial image plane. These transfer coefficients, $\partial L'/\partial N_h$, and $\partial H'/\partial N_h$, which may be termed appropriately the *chromatic coefficients* of the system, provide a means of analysing the importance of the secondary spectrum as it affects the axial and transverse chromatic aberrations of the system. It will also be obvious that they furnish a simple means of determining the effect on the colour correction of changes in the glasses of the system provided these are not too drastic.

The continuation of our example will afford a clear insight into the use of these chromatic coefficients. In table 6 are collected the computed values of the chromatic coefficients for the marginal, zonal, and paraxial rays of the axial pencil, and the principal ray of an oblique pencil at 22°.

Table 6

Comp.	$\partial L'_m/\partial N$	$\partial L'_z/\partial N$	$\partial l'/\partial N$	$\partial H'/\partial N$
a	−737·99	−640·16	−587·86	−39·917
b	768·53	677·23	628·03	30·309
c	−162·67	−157·51	−155·56	− 6·300
d	−127·07	−119·76	−114·53	8·881
e	574·99	502·11	459·53	−30·148
f	−521·13	−446·98	−403·98	38·215

Table 7

Comp.	Chr. Coeff.	Contrbn. b–D	Contrbn. C–D	Contrbn. e–D	Contrbn. –FD	Contrbn. G'–D
a, f	−1087·14	5·6205	3·5658	−2·9788	−8·5340	−15·4591
b, e	1179·33	−6·3566	−4·0333	3·4319	9·8478	18·0320
c, d	− 277·27	1·1257	0·7098	−0·5850	−1·6719	− 3·0195
$L'_z - L'_{zD}$	0·3896	0·2423	−0·1319	−0·3585	− 0·4466	

In table 7 the contributions made by the pairs of components of similar glass to the zonal chromatic aberration, $L'_z - L'_{zD}$, are computed for the various spectral ranges, the last line of the table giving the final value of the aberration. Omitting the corresponding detailed computation for the other three rays under consider-

Table 8

Aberration	b–D	C–D	e–D	F–D	G'–D
$L'_m - L'_{mD}$	0·4443	0·2767	−0·1518	−0·4128	−0·5176
$L'_z - L'_{zD}$	0·3896	0·2423	−0·1319	−0·3585	−0·4460
$l' - l'_D$	0·3634	0·2258	−0·1232	−0·3349	−0·4192
$H' - H'_D$	−0·00255	−0·00158	0·00126	0·00354	0·00638

ation, we collect in table 8 the values of the aberrations calculated for each ray. This provides a sampling of the axial colour correction in the paraxial, zonal, and marginal regions, and a single sampling of the transverse colour at one obliquity.

The curves marked *aa* in figures 4 to 7 show these aberration values plotted against the wave-length, and hence summarize completely the condition of the

system as regards its secondary spectrum. The axial chromatic aberration is very undesirable and calls for attention. A study of the chromatic coefficients and the contributions to the aberrations shows that an improvement should be effected by increasing the V-number of the DBC components and also the HC components.
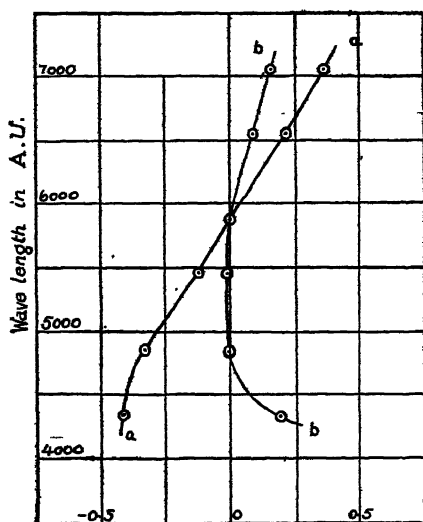
$L'_m - L'_{mD}$

Figure 4.
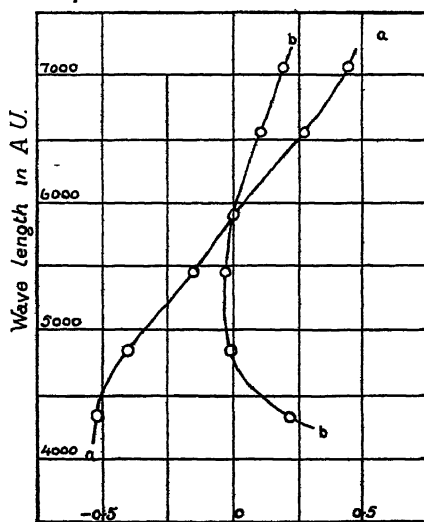
$L_2' - L'_{2D}$
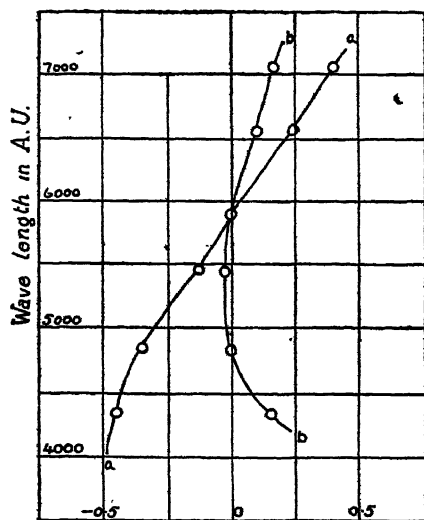
Figure 5.

$l' - l'_D$

Figure 6.
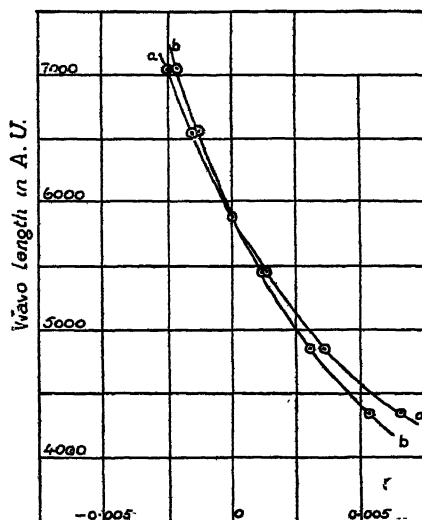
$H' - H'_D$

Figure 7.

By way of example we consider the result of changing to a DBC of the type 613568, and also substitute a BSC of the type 517641 for the HC of components c and d. Ten minutes' computation with the new dispersion values in place of the former ones gives a new set of aberration values. Omitting the details, these are collected

in table 9, from which the curves *bb* in figures 4 to 7 have been drawn. These curves show a very great improvement in the axial colour correction and a slight improvement in the already good transverse aberration. This example affords clear evidence of the general usefulness of these chromatic coefficients of the system.

Table 9

Aberration	b–D	C–D	e–D	F–D	F'–D
$L'_m-L'_{mD}$	0·1888	0·1073	−0·0250	−0·0160	0·2198
$L_z'-L'_{zD}$	0·1627	0·0920	−0·0199	−0·0059	0·2097
$l'-l'_D$	0·1525	0·0861	−0·0196	−0·0069	0·1920
$H'-H'_D$	−0·00221	−0·00136	0·00116	0·00298	0·00523

As a final check on the method we investigate the reliability of the curves *aa* as an analysis of the secondary spectrum. Ray-tracings were made for the three axial rays in colours C and F and the final intersection values for each compared with the values deduced by using the chromatic coefficients.

Table 10

Wave-length	Marginal		Zonal		Paraxial	
	Trace	Coeffs.	Trace	Coeffs.	Trace	Coeffs.
D	108·445	—	106·522	—	106·525	—
C	108·734	108·721	106·766	106·764	106·748	106·751
F	108·030	108·032	106·159	106·163	106·181	106·190

The first line in table 10 gives the intersection lengths for D light for the three rays in the original trace. From these values and the aberration quantities in table 8, the predicted values of the intersection lengths of the rays for C and F light are obtained and entered in the table. The values obtained from the complete traces are entered alongside the predicted values, from which it is seen that the agreement is within the limits of uncertainty of the traced values.

MENDELSSOHN'S Z-PARTICLES AND THE BOSE-EINSTEIN STATISTICS

By WILLIAM BAND,

Yenching University, Peiping, China
(now at Willaston, Wirral, Cheshire)

ABSTRACT. Bose-Einstein statistics are applied to Mendelssohn's *z*-particles and shown to give a satisfactory qualitative explanation of the phenomena of liquid helium.

---

§1. INTRODUCTION

D R. MENDELSSOHN has recently propounded the concept of a *z*-state in which *z*-particles (either atoms or electrons) occupy a set of quantum states lower than the ordinary thermal states and separated from them by a finite energy gap. The *z*-particles form an aggregate in momentum space with zero

thermal energy, and transport phenomena are caused by a diffusion of z-particles under their own zero-point momentum.

As Mendelssohn remarks, this concept should aid the development of a satisfactory theory both of liquid helium and of superconductivity. However, he appears to doubt whether the Bose-Einstein statistics are capable of handling the situation for helium atoms, and favours the possibility that some kind of compromise statistics may be found that will work alike for electrons and helium atoms in the z-state. The present writer believes this to be essentially erroneous, if only because of the fact that the two types of statistics arise from mutually inconsistent symmetry conditions on the wave functions representing the particles ; and it is difficult to see what better compromise can be found than the classical statistics, which obviously do not apply to either case.

The object of the present note is to show that, in fact, Mendelssohn's concept can be harmonized with the Bose-Einstein statistics when the latter is suitably applied to the model proposed by Mendelssohn. Previous attempts to apply the Bose-Einstein statistics to the liquid helium problem have all neglected the possibility that two competing sets of states exist, and that a statistical equilibrium has to be set up between them. When this equilibrium is set up, the Bose-Einstein statistics give a very satisfactory explanation of the phenomena in liquid helium.

§2. STATISTICAL CHARACTERISTICS OF THE Z-STATE

The z-particles with zero, or at any rate with less than normal, thermal energy must be treated statistically as having fewer degrees of freedom than the three normally characterizing the thermal states. If it is provisionally assumed that they have but two degrees of freedom, we at once arrive at a suggestion of a mechanism whereby this reduction may in fact occur : because the z-state is certainly associated very intimately with the *surface* layers of liquid helium, especially in the frictionless transport of helium atoms in films.

Without, however, attempting to specify the mechanism, let us accept the hypothesis that the z-particles have two degrees of freedom only, and that as a crude approximation we may regard the energy states of the z-particles as determined by Schrödinger waves in resonance with the surface of the assembly. This concept agrees qualitatively with Mendelssohn's requirement that the z-particles shall have a higher degree of order than the thermal states, and form a separate phase in momentum space alone. It is further necessary to assume, so as to coincide with the Mendelssohn model, that the energy zero of the surface states is at a finite level below the lowest thermal states, so that there will be no momentum transfer from the surface or z-states at sufficiently low temperature.

The energies of the thermal states are

$$\epsilon_t = h^2(s^2+t^2+u^2)/8m_t V^{2/3}, \quad s, t, u = 1, 2, \ldots \qquad \ldots\ldots(1)$$

and of the z-states

$$\epsilon_z = -\chi + h^2(s^2+t^2)/8m_z A, \quad s, t = 1, 2, \ldots \qquad \ldots\ldots(2)$$

where A is a crude approximation for the effective area of the assembly, and $-\chi$ is the Mendelssohn energy gap: by hypothesis χ is positive.

§3. STATISTICAL EQUILIBRIUM BETWEEN Z-STATES AND THERMAL STATES

If there are in all N particles distributed among the two sets of states, then the Bose-Einstein statistics will lead to the equilibrium given by

$$N = N_t + N_z, \qquad \ldots\ldots(3)$$

where

$$N_t = \sum_{stu} \sum_{j=1}^{\infty} \lambda^j \exp\left[-\frac{jh^2(s^2+t^2+u^2)}{8m_tkTV^{2/3}}\right] \qquad \ldots\ldots(4)$$

and

$$N_z = \sum_{st} \sum_{j=1}^{\infty} \lambda^j \exp\left[j\chi/kT - \frac{jh^2(s^2+t^2)}{8m_zkTA}\right] \qquad \ldots\ldots(5)$$

are the mean numbers in the thermal and z-states respectively.

Write

$$\lambda^* = \lambda e^{\chi/kT} > \lambda, \qquad \ldots\ldots(6)$$

so that

$$N_z = \sum_{st} \sum_{j=1}^{\infty} \lambda^{*j} \exp\left[-\frac{jh^2(s^2+t^2)}{8m_zkTA}\right]. \qquad \ldots\ldots(7)$$

The expression (4) was studied by Fowler and Jones. It was summed by means of a theta-function transformation and shown to approach the following limit for large N:

$$N_t = (2\pi m_t kT)^{3/2} V h^{-3} \sum_{j=1}^{\infty} \lambda^j/j^{3/2}, \qquad \ldots\ldots(8)$$

provided

$$T > T_0, \qquad \lambda < 1,$$

where

$$2\pi m_t kT_v \sum_{j=1}^{\infty} 1/j^{3/2} = (N_t/V)^{3/2} h^2. \qquad \ldots\ldots(9)$$

The corresponding summation of (5) leads to

$$N_z = 2\pi m_z kTA h^{-2} \sum_{j=1}^{\infty} \lambda^{*j}/j. \qquad \ldots\ldots(10)$$

So long as $\lambda^* < 1$ the number $N_z \ll N_t$ and can be neglected. The parameter λ is determined then by the equation

$$N = (2\pi m_t kT)^{3/2} V h^{-3} \sum_{j=1}^{\infty} \lambda^j/j^{3/2}.$$

As T falls, λ increases, and as soon as

$$\lambda \to e^{\chi/kT} < 1, \qquad \lambda^* \to 1,$$

the series for N_z will begin to diverge, so that N_z can no longer be neglected. There is thus a critical temperature, where

$$N = (2\pi m_t kT_c)^{3/2} V h^{-3} \sum_{j=1}^{\infty} e^{-j\chi/kT_c}/j^{3/2}. \qquad \ldots\ldots(11)$$

Below this temperature λ remains almost at the value $e^{-\chi/kT}$, so (3) becomes

$$T \leqslant T_c$$

$$N = (2\pi m_t kT)^{3/2} V h^{-3} \sum_{j=1}^{\infty} e^{-j\chi/kT}/j^{3/2} + 2\pi m_z kTA h^{-2} \sum_{j=1}^{\infty} \lambda^{*j}/j. \qquad \ldots\ldots(12)$$

The population of the z-states is thus negligible for $T > T_c$, and below T_c it is given by

$$N_z = N\left[1 - \left(\frac{T}{T_c}\right)^{3/2}\left\{\frac{\sum_{j=1}^{\infty} e^{-j\chi/kT}/j^{3/2}}{\sum_{j=1}^{\infty} e^{-j\chi/kT_c}/j^{3/2}}\right\}\right]. \qquad \ldots\ldots(13)$$

As Mendelssohn remarks, the variation of N_z is more rapid than would be given by

$$N_z = N \left[1 - (T/T_c)^{3/2}\right],$$

and (13) will in fact give good agreement with experimental data provided χ is suitably chosen.

§4. CONCLUDING REMARKS

The present writer, with Dr. Hsu, formerly (1941) worked out a theory of liquid helium based upon the idea of surface resonance, and obtained remarkable agreement with observational data on specific heat and thermomechanical phenomena. The chief weakness of that theory was that no definite explanation was found for the frictionless nature of the film transport, although indeed the concept of diffusion by zero-point momentum (there called intrinsic pressure) was suggested.

Mendelssohn has now pointed out that a finite energy gap between the z-states and the thermal states will prevent frictional transfer, and by incorporating this idea, as in the present paper, the Bose-Einstein statistics can certainly be made to give a satisfactory qualitative explanation of the phenomena of liquid helium.

There are two important details which the theory cannot handle in its present form: First, the critical temperature is in fact hardly more than half that predicted by (11), and second, the critical temperature follows the λ-line in its dependence on external pressure.

But it is almost obvious that the discrepancy here is due to molecular forces, and the fact that the energy states (1) and (2) are derived on the assumption that the container is free from potential irregularities. It is clearly quite hopeless to improve matters in this respect until a better theory of liquids is available, and the Bose-Einstein statistics applied to the helium system as a liquid rather than a gas.

REFERENCES

BAND, W. and HSU, H. Y., 1941. *Phys. Rev.*, **59**, 1016.
FOWLER, R. S. and JONES, H., 1938. *Proc. Camb. Phil. Soc.*, **34**, 573.
MENDELSSOHN, K., 1945. *Proc. Phys. Soc.*, **57**, 371.

DISCUSSION

on paper by S. BAXTER entitled " The thermal conductivity of metals "
(*Proc. Phys. Soc.*, **58**, 105).

Mr. J. B. CARNE. I am prompted to draw attention to the account of the early work of Sir Benjamin Thomson, which was presented to the Royal Society in 1786 *et seq.*

From a systematic series of beautifully simple experiments, Sir Benjamin deduced not only the now generally accepted explanation of the mechanism of heat transfer by convection, but also the existence of a layer of stagnant air which acts as a barrier to the transmission of heat from a hot surface, the postulate invariably attributed to Langmuir.

Reference to the papers will also show that Thomson found the same variation of conductivity with density for very low bulk densities of air-fibre mixtures as recorded by Baxter. It is interesting to note also that Sir Benjamin Thomson also enquired into the relation between water content of air-fibre mixtures and their conductivity.

THE CONSTRUCTION AND USE OF A "FLY'S EYE" FOR ASSISTING X-RAY STRUCTURE ANALYSIS

By ALEXANDER R. STOKES,

Cavendish Laboratory, Cambridge

(now at Royal Holloway College, Englefield Green, Surrey)

ABSTRACT. The "fly's eye" is the name given to a multiple camera for producing two-dimensional diffraction gratings consisting of repeated patterns representing projections of crystal structures. This paper describes the construction of a fly's eye using lenses embossed on Perspex ; also the necessary adjustments, the method of photographing a structure, and the apparatus for viewing the spectra produced by the two-dimensional gratings so obtained.

§ 1. INTRODUCTION

THE purpose of the "fly's eye" is to produce on a photographic plate a regularly repeated pattern representing the projection along some crystallographic direction of a crystal structure. This behaves as a two-dimensional diffraction grating and gives spectra whose intensities are proportional to the intensities of x-ray reflections in a zone corresponding to the direction of projection. The use of the fly's eye thus obviates the calculation of structure amplitudes of a proposed structure in the trial-and-error stages of structure determination. The purpose of the present paper is to describe the construction and setting-up of a fly's eye; the theory of its use is to be given in subsequent papers.

§ 2. DEVELOPMENT OF THE INSTRUMENT

The first method by which a repeated pattern was produced (Bragg, 1944) was by means of a pin-hole camera with an array of pin-holes lying on a square lattice. This was made by photographing an array of black dots on a white background, giving an opaque plate with an array of transparent holes. These were spaced at 0·025 cm., each hole having a diameter of 0·0012 cm., and were arranged on a square with 26 holes along each side. To produce a photograph representing a structure a photographic plate was supported 1 mm. above the pin-holes, and a light, which could be moved about in a horizontal plane 80 cm. below the plate, was switched on for a short time in each of the several positions representing atomic positions in the unit cell. The pin-hole camera gave a reduction in linear dimensions of 800 : 1, and the size of the "unit cell" on which the lamp was moved was, therefore, 0·025 × 800 = 20 cm.

This method gave good results but suffered from two disadvantages: the pin-hole images are somewhat diffuse, and specks of dust may block the pin-holes and give rise to images of unequal intensities.

It is therefore advantageous to use lenses rather than pin-holes. Small lenses can be made in the following way: if a steel ball-bearing is pressed into a copper

plate, a spherical depression of nearly optical perfection is produced; if, now, a piece of Perspex is pressed at its softening temperature into the copper plate, a spherical lens will be embossed on the surface of the Perspex.

To make a fly's eye we require a regular array of such lenses, and we want, therefore, to make a regular array of depressions in a copper plate. Two ways of obtaining such an array are (1) by packing a large number of balls into a tray and pressing a copper plate on to them, (2) by making each depression individually and moving the copper plate through the repeat distance of the pattern between successive depressions.

The first method suffers from one inherent disadvantage, namely, that it fixes the spacing between the depressions relative to their radius of curvature. The distance between adjacent balls in the tray must be twice their radius, r, and since the lenses will have a focal length of $r/(\mu-1)$, this distance is necessarily $2(\mu-1)$
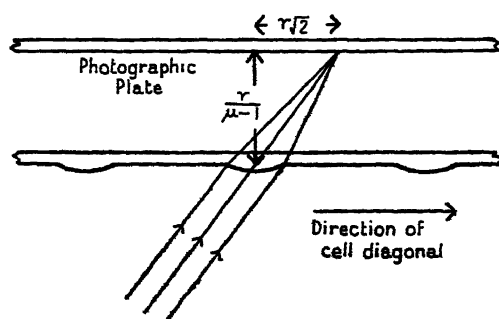
Figure 1. Diagram showing maximum inclination of rays to axes of lenses.

times the focal length of the lenses, where μ is the refractive index of Perspex, equal to 1·49. Therefore rays which would come to a focus at the *corner* of a unit cell (see figure 1) would have to be inclined to the axis at an angle of

$$\tan^{-1}(\mu-1)\sqrt{2}=35°.$$

This is too large an angle between rays and axis for a plano-convex lens of reasonable aperture, and it cannot be reduced except by making $(\mu-1)$ smaller; this could be done by placing another transparent medium over the lenses, but this would introduce complications in construction.

The second method has the disadvantage that the spacing and diameter of the lenses are not so easily kept constant, but with suitable precautions it was found possible to make the variations in these unimportant.

Construction *

A cylindrical piece of copper 1·5 in. in diameter and 0·5 in. thick was polished on one of its plane faces, first with emery paper and then with metal polish. In order to provide an accurately measurable translatory motion, the copper was screwed firmly, with its polished side upwards, to a lathe tool-rest with two gradu-

* All dimensions are given in inches. since machine tools and ball-bearing balls are calibrated in inches.

ated screw-motions at right angles, which in turn was screwed to the base plate of a drilling machine. A steel ball, 3/16 in. in diameter, was fixed with wax into a depression in the end of a steel rod, which was supported in the drill holder of the machine. A weight was suspended from the arm of the drilling machine and, by lowering this weight slowly, the ball was brought down with a constant thrust (as in a hardness-testing machine) on the copper plate. (The usual rotary motion of the drilling machine was not, of course, used.) The weight was adjusted by trial so that a depression of 0·025 in. diameter was made. The plate was moved 0·05 in. after each depression was made. A slight "burr" was thrown up around the depressions, and it was therefore important not to make them too large in relation to their distance apart.

The depressions were made to lie on a square lattice with a spacing of 0·05 in. and to cover a circular area 1·05 in. in diameter; there were 341 depressions altogether. Their diameters were constant to about 10%, which was sufficiently good for this purpose.

The lenses were then made from this negative by pressing a piece of Perspex between the copper plate and a polished brass plate in a moulding press. The Perspex was originally 0·15 in. thick and was cut to a circle 1·5 in. in diameter. The moulding press was heated to about 140° c., at which temperature the Perspex softened, and was screwed down slightly; about 0·02 in. was sufficient to emboss the lenses. The pressure was not released until the temperature had fallen below 80° c., so that the Perspex was hard again.

The lenses so produced had a numerical aperture of about $f/7·5$, and were found by photographic tests to give good definition over the whole field of view employed.

In order that light should pass only through the lenses, it was necessary to make the plane parts of the Perspex opaque. Methods of inking or painting over the plane parts were tried but were found impracticable, as no ink or paint could be found which was opaque in layers as thin as the embossed lenses. Instead, a mask was made by punching holes of the same diameter as the lenses in a piece of black paper. A needle with its end ground flat was used as a punch and the paper was supported on a flat piece of lead. The drilling machine and lathe tool-rest were used, as described above, to place the holes correctly. The mask was stuck to the Perspex with "Durofix", which was applied only round the edges of the fly's eye, to avoid contaminating any of the lenses.

Setting up

In order that a photographic plate may be supported in the focal plane of the lenses, three holes were drilled and tapped near the circumference of the Perspex disc, and three 8-B.A. screws, pointed at their tips, were inserted in these holes and provided with nuts for locking. The fly's eye was fixed by these screws on to a circular brass plate with a hole in its centre large enough to include all the lenses. This was mounted on a horizontal shelf of adjustable height with the lenses downwards and the tips of the screws upwards (figure 2). Below this was a base plate on which a lamp could be moved to various known positions by reference to two scales at right angles. This could be replaced when desired (see below) by an

illuminated ground-glass screen. The whole apparatus must, of course, be used in a darkened room.
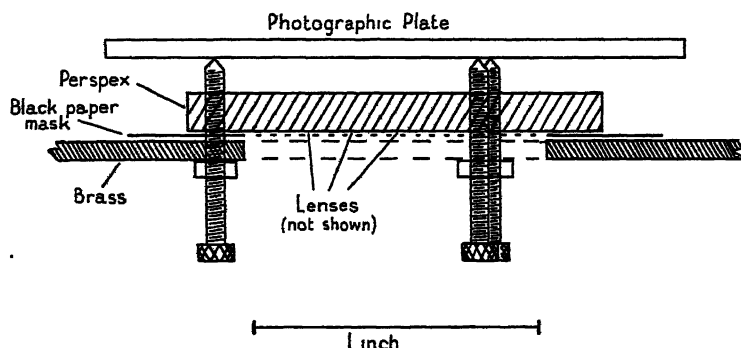
Figure 2. Set-up of fly's eye.

Adjustments

(i) *Focus.* The three screws were adjusted, before the fly's eye was finally set up, by placing the fly's eye on a microscope stage with a piece of ground glass resting on the screws, and turning the screws so that all the images of a distant object were seen in the same plane as the ground glass. A blue filter was used so as to obtain the focus for blue light.

A slight adjustment of the focus was now made to allow for the fact that the photographic plate responded mainly to ultra-violet light. This was done by lowering one of the screws by about half a turn and taking a fly's eye photograph of the lamp in a number of positions. Since the photographic plate is tilted during
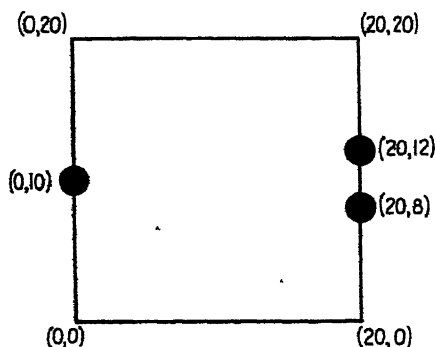
Figure 3. Unit cell and positions of lamp.

this exposure, the images are not in focus all over the plate; by finding those images for which the definition is best, the height of the focal plane could be calculated, and hence the screws were adjusted to the correct height.

(ii) *Scale and orientation.* Two adjustments are now necessary to ensure that the contents of the unit cell shall be reproduced in the fly's-eye photograph at the correct scale and in the correct orientation. This was done as follows: a photograph was taken for which the lamp was switched on in positions whose coordinates were (0,10), (20,8), (20,12), measured in cm., the "unit cell" being 20×20 cm. (figure 3). The fly's eye should, if correctly adjusted, give a pattern as

shown in figure 4 (*a*), where the three black circles represent the spots belonging to one unit cell. A typical out-of-adjustment photograph is represented by figure 4 (*b*),
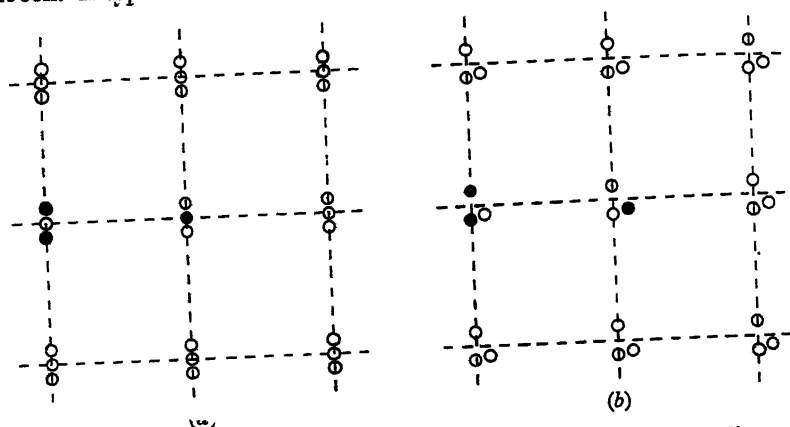
Figure 4. Photographs obtained with (*a*) correct adjustment, (*b*) incorrect adjustment.

which shows that a rotation of the fly's eye and a reduction of scale are necessary. To reduce the scale, the shelf on which the fly's eye is mounted must be raised.

§3. METHOD OF PHOTOGRAPHING A PROPOSED STRUCTURE

The method used with the original pin-hole fly's eye was to switch on the lamp for a suitable length of time in each of the positions at which an atom was to be represented, so that each atom was reproduced as a black spot in each unit cell of the fly's-eye photograph. A negative of this photograph, in which each atom would be represented by a white spot, was then taken, and diffraction spectra from it were observed.

This method may also be used for the lens fly's eye, but a more convenient method is available. If the lamp is replaced by a uniformly illuminated screen, 20 cm. square, and a number of opaque discs are placed on the screen, the fly's-eye photograph will have a black background with light circles on it representing atoms. This method has the advantages that no negative need be taken, and that only one exposure is needed, once the discs have been arranged on the screen, instead of one for each atom.

In practice, a difficulty arises here because the lenses of the fly's eye may not all be accurately spaced, and, therefore, the black squares which are the images of the light square 20 × 20 cm. may overlap in some places and fail to meet in others. Overlapping does not matter provided the plate is sufficiently over-exposed for all the black part to be practically opaque: the failure of some of the squares to meet may be overcome by making the illuminated squares slightly larger than 20 × 20 cm., so that all the squares overlap a little. Regions near the edge of the unit cell in the fly's-eye image of the screen are now illuminated twice over, by light from opposite edges of the screen, and no parts are left unilluminated. Any atom that lies on or near the edge of the unit cell must now be represented twice over, as shown at AA in figure 5 (*a*), to ensure that the image of A shall be a complete shadow. Similarly an atom near the corner of the cell must be represented four times over, as shown in figure 5 (*b*).

(a)   (b)
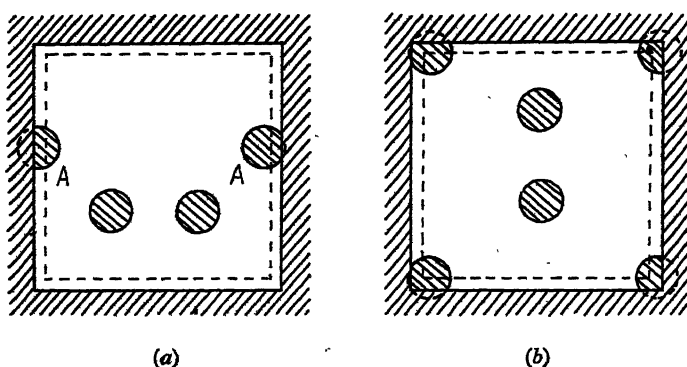
Figure 5.   Repetition of atom near edge or corner of unit cell.

§4. THE DIFFRACTION APPARATUS

The apparatus for observing the diffraction spectra was similar to that already described by Bragg (1944). It consisted essentially of a mercury-vapour arc lamp with a pin-hole stop, a lens of about a metre focal length to produce a collimated beam, in which the two-dimensional grating was placed, and a similar lens to focus the spectra. Brass tubes were arranged to eliminate extraneous light. The separation of the spectra thus produced was only about 0·3 mm.; an eyepiece or low-power microscope was therefore used for observing the spectra. A green or deep blue filter was found useful for separating a single line of the mercury spectrum; this was almost essential when the spectra were to be photographed.

As the emulsion and the glass of the grating are not of very uniform thickness, the diffraction pattern may be improved considerably by placing the plate between two optical flats and filling the spaces with cedarwood oil.

§5. TEST OF FLY'S EYE ON DURENE STRUCTURE

The fly's eye was tried out on the known structure of durene (Robertson, 1933). The unit cell has not a square cross-section, but if all the atomic-position coordinates are referred to equal axes at right angles the intensities of diffraction are not altered. The projection of the structure along the b-axis was taken, and since this gives two identical parts to each unit cell, the cell was halved before referring it to equal axes at right angles. Each carbon atom was represented by a disc laid on an illuminated screen as described above, while the hydrogen atoms were neglected. An enlarged reproduction of the fly's-eye photograph is shown in figure 6, while in figures 7 (a) and 7 (b) are shown the diffraction photographs obtained from this: the first was taken with the fly's-eye photograph enclosed between optical flats and cedarwood oil, while the second was taken without the use of optical flats. The figures under the spots in figure 7 (a) are relative intensities calculated from Robertson's data. The agreement is seen to be good when allowance is made for the fact that the discs do not properly represent the carbon atoms, giving a diffracted intensity which does not fall off rapidly enough with angle; also the effect of temperature on intensity is neglected.
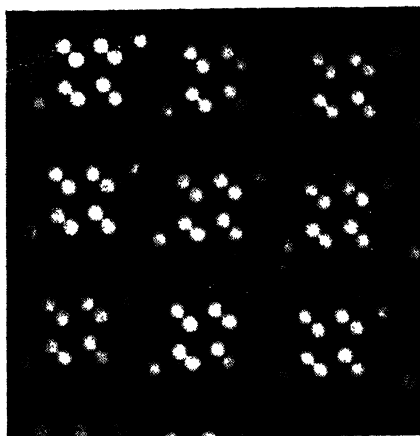
Figure 6.   Fly's-eye photograph of durene.

(a)   (b)
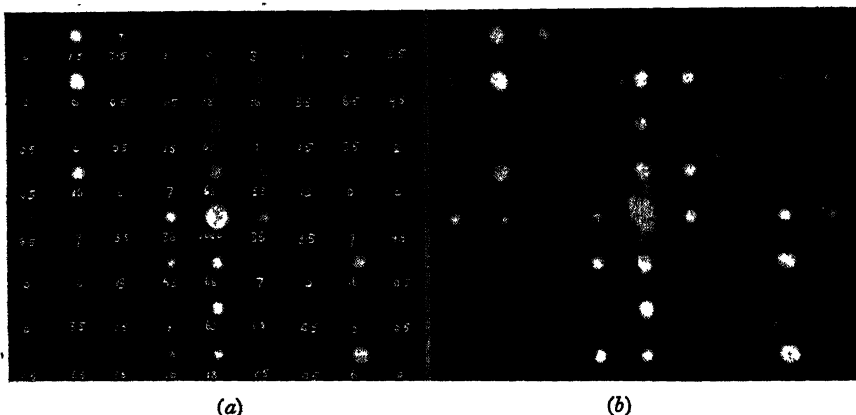
Figure 7.   Diffraction pattern from durene grating : (a) optical flats and cedarwood oil used ;
(b) no optical flats used.

ACKNOWLEDGMENTS

I wish to thank Prof. Sir Lawrence Bragg, F.R.S., under whose guidance the work was started, and Dr. H. Lipson and Dr. G. L. Rogers for many helpful discussions.   I am indebted also to the Master and Fellows of Trinity College for financial support through a Senior Rouse Ball Studentship.

REFERENCES

BRAGG, W. L., 1944.   *Nature, Lond.*, 154, 69.
ROBERTSON, J. M., 1933.   *Proc. Roy. Soc.*, A, 142, 659.

DISCUSSION

Mr. D. E. THOMAS.   Attempts to follow Dr. Stokes and Prof. Bragg's technique have been made in my laboratory at the Research Department, Woolwich, and one or two minor differences have arisen.   The original mould was made with a Vickers hardness-testing machine, the brass block being held in a clamp which could be traversed in two directions at right angles by means of micrometer screws.   The technique of making the mask differed considerably.   An image of the Perspex fly's eye was made by focusing a

distant small source of light on to a lantern slide. The image was intensified by multiple printing until it gave clear holes on a black background. It was then fitted immediately over the fly's eye.

After forming the diffraction grating it was reduced to one-third (linear dimensions) on a fine resolution plate, and this small grating used to give the diffraction patterns of a point source of monochromatic light. These patterns can be seen with the naked eye and photographed, and are of a convenient size for records.

Dr. R. S. CLAY. I suggest that the author could have made the impressions more easily if he had used the group of balls in the frame, and obtained closer grouping of the impressions by moving the frame, after making one impression, an exact fraction of the distance apart of the balls (e.g. a third of the distance apart), then making a second impression, and so on. With nine impressions (in the case assumed) he would get a complete mould.

Prof. A. O. RANKINE. Dr. Stokes's paper emphasizes a certain ambiguity in the use of the word "spectra" as between x-ray and ordinary light phenomena. It sounds a little odd to hear of a blue filter being inserted to render the spectra more definite, although the meaning, on contemplation, is evident enough. Must diffraction patterns originating from single wave-length radiation on the one hand, and arrays of images produced by composite wave-length radiation on the other, both be designated by the same term—spectra?

AUTHOR'S reply. I am interested in Mr. Thomas's method of making and using his fly's eye. With regard to Dr. Clay's suggestion, I doubt whether the advantage of this method would outweigh the trouble of both fitting balls in a frame and providing the necessary translatory motion ; moreover, a much greater force would be needed in order to produce many indentations at once. It is, incidentally, somewhat difficult to persuade balls to form a square array in a tray ; they always tend to go into a hexagonal pattern.

In answer to Prof. Rankine's point about spectra, my view is that, in x-ray or optical diffraction, one spectrum appears on a photographic plate generally as a series of spots, one for each wave-length of the incident waves ; if these be monochromatic, then a spectrum consists of one spot only. The several spots which arise from the different orders of diffraction may thus be correctly called spectra. It is, however, more precise, though less convenient, to speak of " orders of diffraction " rather than " spectra ".

FLUCTUATION NOISE IN A RECEIVING AERIAL

By R. E. BURGESS,

National Physical Laboratory, Teddington

ABSTRACT. The factors which determine the signal/noise ratio at the terminals of a receiving aerial are discussed. The aerial noise considered is the random fluctuation type, consisting of

    (i) thermal noise associated with the loss resistance of the aerial, and

    (ii) noise associated with the radiation resistance which is induced by the surroundings.

The effective temperature of the radiation resistance is expressed in terms of the temperature distribution of the surroundings and the distribution of power dissipation when the aerial is transmitting. Radiation from the sun and from the Milky Way are briefly discussed, and it is shown that the detection of solar radiation at radio frequencies requires the use of highly directional aerials.

The limitations imposed by the receiver noise on
- (a) the sensitivity of the reception of signals, and
- (b) the accuracy of measurement of aerial noise, are discussed and the results presented graphically.

The conclusions are of most practical interest at the higher radio frequencies (above about 20 Mc./s.) where atmospheric noise is negligible.

---

§ 1. INTRODUCTION

IT has been shown (Burgess, 1941) that for an aerial at thermal equilibrium in an enclosure at temperature T, the mean square noise e.m.f. has the value given by Nyquist's theorem,

$$e_n{}^2 = 4kTRB = 4kT(R_0 + R_r)B, \qquad \ldots\ldots(1)$$

where R is the total resistance and B is the band-width of the receiving system. This noise is composed partly of the intrinsic thermal agitation e.m.f. generated in the loss resistance R_0 of the aerial and partly of the noise received from the surroundings which can be attributed to the radiation resistance R_r.

In practical aerial systems these ideal conditions do not obtain, and the effective temperatures T_0 and T_r of the ohmic and radiation resistances will not in general be equal. The mean square noise e.m.f. is thus

$$e_n{}^2 = 4k(T_0 R_0 + T_r R_r)B. \qquad \ldots\ldots(2)$$

It is assumed throughout that all the noise is of the random fluctuation type, in which the mean square e.m.f. is strictly proportional to band-width.

The present note is supplementary to the earlier paper, and discusses the influence of the aerial characteristics and the receiver noise on the limiting sensitivity of signal reception and on the detection and accuracy of measurement of noise fields.

§ 2. THE EQUIVALENT CIRCUIT OF A RECEIVING AERIAL FOR SIGNAL AND NOISE

The effective height of an aerial for reception is a function of the direction of arrival and polarization of the received wave. If the wave has electric intensity E at some reference point in the aerial system, the induced signal e.m.f. appearing at the terminals is given by

$$\cdot e_s = Eh(\theta, \phi, \alpha), \qquad \ldots\ldots(3)$$

where h is a function of the direction of arrival (θ, ϕ) and polarization (α) as well as of frequency and of the choice of terminals.

The effective height for transmission has the same value and the radiation resistance of the aerial can be found readily from Poynting's theorem:

$$R_r = \frac{30}{\lambda^2} \int_0^{2\pi} d\alpha \int h^2 . d\Omega, \qquad \ldots\ldots(4)$$

where λ is the wave-length, while $d\Omega = \sin\theta . d\theta . d\phi$ is the element of solid angle. The equation gives R_r in ohms if λ and h are in metres; M.K.S. units are assumed throughout this paper. It is not customary to take account of the dependence of effective height on polarization, but this is essential in the present analysis

since, in the transmitting condition, an aerial in general produces elliptical polarization at a distant point. Furthermore, the random polarization of noise fields requires integration over all angles of polarization in order to calculate the induced noise e.m.f. in the case of reception.

The directivity or gain of an aerial is a measure of the extent to which its polar diagram is concentrated in one direction and for one polarization. Thus if h_m is the maximum effective height which occurs for this direction and polarization, the directivity can be defined by

$$D = \frac{4\pi^2 h_m^2}{\int_0^{2\pi} d\alpha \int h^2 . d\Omega} . \qquad \dots\dots(5)$$

The directivity is unity for a hypothetical omnidirectional aerial giving linear polarization ($h = h_m \sin\alpha$). For linearly polarized aerials the above equation agrees with the usually accepted definition of directivity, e.g. for a Hertzian dipole $D = 1\cdot 5$ and for a half-wave dipole $D = 1\cdot 64$.

Combining equations (4) and (5) gives

$$R_r = \frac{120\pi^2 h_m^2}{\lambda^2 D} \text{ (ohms).} \qquad \dots\dots(6)$$

The *maximum* available power for such an aerial in a plane wave of electric intensity E volts/metre is

$$P = \frac{E^2 h_m^2}{4R_r} = \frac{E^2 \lambda^2 D}{480\pi^2} \text{ watts}$$
$$= F\lambda^2 D/4\pi \text{ watts,} \qquad \dots\dots(7)$$

where

$$F = \frac{E^2}{120\pi} \text{ watts/m}^2$$

is the energy flux in the wave.

In practice the resistance R of an aerial is the sum of the radiation resistance R_r and the resistance R_0, which represents the losses in the conductors, the dielectrics and the earth (and in the terminating resistance in the case of a progressive wave system), all at the ambient temperature T_0. It is convenient to write for the ratio of the radiation resistance to the total resistance

$$y = \frac{R_r}{R},$$

where y may for obvious reasons be termed the efficiency. For a resonant dipole y may closely approach unity, but for a short aerial or a progressive wave-system it may be small.

Now the effective noise temperature of a linear two-terminal system whose elements are at different temperatures T_k is readily shown by an application of the Reciprocity Theorem to be given by

$$T' = \frac{\Sigma P_k T_k}{\Sigma P_k} = \frac{\Sigma P_k T_k}{P}, \qquad \dots\dots(8)$$

where P_k is the power dissipated in the kth element when a generator is applied to the terminals of the system and P is the total power dissipated.

This equation is directly applicable to an aerial. The loss resistance R_0, which corresponds to the power dissipated in the aerial itself and in its immediate surroundings, is at the ambient temperature T_0. The radiation resistance R_r corresponds to the power which will eventually be dissipated, at a relatively great distance from the aerial, and thus its effective temperature cannot be so readily determined. For example, on low radio frequencies (say below 1 Mc./s.), radiation can neither escape from nor enter the space bounded by the earth and the ionosphere, and in this case the effective temperature T_r of the radiation resistance will lie somewhere between the temperature of these boundaries. At higher radio frequencies, where penetration of the ionosphere occurs, the temperature of bodies lying outside it must be taken into account.

The equivalent circuit of the receiving aerial for signal and noise may now be readily constructed. The signal e.m.f. is hE, where h is the effective height for the particular conditions of the incident wave, and will be h_m if adjustment is made for maximum reception. The noise e.m.f.s are $\sqrt{(4kT_0R_0B)}$ in series with the ohmic resistance R_0 and $\sqrt{(4kT_rR_rB)}$ in series with the radiation resistance R_r. The noise e.m.f.s may be combined by ascribing an effective temperature T' to the whole aerial resistance R, where

$$T' = \frac{R_rT_r + R_0T_0}{R} = yT_r + (1-y)T_0. \qquad \ldots\ldots(9)$$

The overall signal/noise ratio at the aerial terminals in the condition of maximum reception is thus

$$\frac{e_s}{e_n} = \frac{h_mE}{\sqrt{4kBR_r\left(T_r + \frac{1-y}{y}T_0\right)}}$$

$$= \frac{\lambda E}{\sqrt{480\pi^2kBT_r}} \sqrt{\frac{D}{1 + \frac{1-y}{y}\frac{T_0}{T_r}}}. \qquad \ldots\ldots(10)$$

The first factor represents the signal/noise ratio of a perfectly efficient aerial ($y=1$) of directivity unity. To render practical interpretation easy, figure 1 shows the field strength for which the signal/noise ratio is unity in a loss-free Hertzian dipole ($D=1\cdot5$, $y=1$) with a band-width B of 10 kc./s. and an ambient temperature of $290°\mathrm{K}$. The corresponding equation is

$$E_1 = \frac{\sqrt{(320\pi^2kBT_r)}}{\lambda} = \frac{0\cdot35}{\lambda}\sqrt{\frac{T_0}{T_r}} \qquad \ldots\ldots(11)$$

(μv./m. when the wave-length is in metres).

It is seen from equation (10) that if the noise field is relatively large ($T_r \gg T_0$) the efficiency of the aerial is of secondary importance. It is on account of the prevalence of high noise levels that aerial systems, which theoretically appear to be poor noise collectors of signal energy, are in practice found to be little inferior to systems designed to be of high efficiency.

If the noise radiation is not perfectly diffuse, T_r will itself be a function of the directivity and, in fact, may either increase or decrease as D increases, depending upon whether the polar diagram is concentrated on a region of relatively higher or lower temperature.

§3. THE EFFECTIVE TEMPERATURE OF THE RADIATION RESISTANCE

The density of black-body radiant energy at absolute temperature T in a frequency band B is given by the Rayleigh-Jeans law at radio frequencies:

$$U = \frac{8\pi kT}{c\lambda^2} B.$$

The energy flux received from a solid angle $d\Omega$ is given by

$$dF = \frac{dE_n^2}{120\pi} = \frac{cU}{4\pi}.d\Omega = \frac{2kTB}{\lambda^2}.d\Omega \text{ watt/m}^2,$$

where dE_n^2 is the element of mean square noise field intensity.* Thus the total mean square noise e.m.f. induced in an aerial from its surroundings is given by

$$e_n^2 = \frac{240\pi kB}{\lambda^2} \int_0^{2\pi}\frac{d\alpha}{2\pi}\int Th^2.d\Omega.$$

Therefore the effective temperature of the radiation resistance is given by

$$T_r = \frac{e_n^2}{4kBR_r} = \frac{\displaystyle\int_0^{2\pi} d\alpha \int Th^2.d\Omega}{\displaystyle\int_0^{2\pi} d\alpha \int h^2.d\Omega} \qquad \dots\dots(12)$$

by virtue of equations (5) and (6). This formula for T_r is seen to be formally identical with equation (8) for a linear network, and, in fact, $h^2.d\Omega.d\alpha$ is proportional to the element of power dissipated in a direction $d\Omega$ for angles of polarization between α and $(\alpha + d\alpha)$.

If the source of radiation is a black-body at temperature T which subtends only a small solid angle Ω at the aerial, and furthermore, if the aerial is linearly polarized and directed towards the body, equation (12) assumes the simple form

$$T_r = T\frac{\Omega D}{4\pi}, \qquad \dots\dots(13)$$

provided that the aerial directivity D is small compared with $\frac{4\pi}{\Omega}$. To take a specific instance, consider the case of black-body radiation from the sun for which approximately

$$T = 6000° \text{K.}, \quad \Omega = 7 \times 10^{-5},$$
$$F = 1\cdot12 \times 10^{-13}/\lambda^2 \text{ watts/m}^2 \text{ per Mc./s.,}$$
$$T_r = 0\cdot03D°\text{K.,}$$

and thus it will be appreciated that very high directivity of the order of 10^4 is required if the solar radiation is to contribute appreciably to the aerial noise. Furthermore, it is not certain that the sun will radiate as a black-body or that absorption in the intervening space will be negligible, and it is, therefore, possible that the noise field at the aerial will differ from the ideal case calculated above. Measurements of noise at radio frequencies of about 20 Mc./s. by Jansky (1935

---

* dE_n^2 is *twice* the mean square field measured in any one direction in the wave-front, the factor of 2 taking into account the two independent orthogonal polarizations which contribute equally to the energy flux of a randomly polarized wave.

and 1937) did not reveal any evidence of solar radiation, nor did those of Reber (1940 and 1942) at higher frequencies (160, 900 and 3300 Mc./s.).*

Jansky (1935 and 1937), Reber (1940 and 1942), and Franz (1942) have established that an intense source of fluctuation noise exists in a fixed region in space in the direction of the Milky Way. The exact nature and origin of this radiation has not yet been conclusively established, but if it is thermal, the temperatures at the source must be very high (probably greater than 100,000°K.). The discussion by Henyey and Keenan (1940) is of interest in this connection.

It is suggested that refined measurements of noise radiation at relatively high frequencies for which the ionosphere and the earth's atmosphere are sensibly transparent, and at which directive aerials are readily constructed, would provide much useful information regarding the temperature of, and processes in, regions of outer space

§4. THE SIGNIFICANCE OF RECEIVER NOISE

If an aerial is attached to an ideal receiver which produces no noise, the output signal/noise ratio (measured at a point prior to detection) would be the same as that at the aerial. Similarly any source of noise radiation whose field acts on the aerial could be detected provided that sufficient amplification were available.

In practice a receiver contributes to the output noise level, so resulting in a reduction of signal/noise ratio in the case of signal reception or in masking of the aerial noise where measurement of the latter is desired.

A convenient measure of receiver noise is the " noise factor" usually denoted by the symbol N. A receiver of noise factor N connected to a source of signal, whose resistive component R of impedance is at temperature T_0, will give a signal/noise ratio which is $1/N$ times that given by an ideal noise-free receiver of the same band-width connected to the same source. The noise e.m.f. of the receiver itself referred to the aerial terminals is, therefore,

$$e_r^2 = 4kT_0(N-1)RB. \qquad \qquad \ldots \ldots (14)$$

(In practice it is more convenient to express the noise factor as $10 \log_{10} N$ decibels, but in this equation and those that follow the numerical power ratio is used.) It is seen from this equation that the noise of the receiver can be expressed as an equivalent field intensity by taking $(N-1)$ for T_r/T_0 in figure 1.

The loss in signal/noise ratio when the receiver is joined to an aerial at effective temperature T' will not in general be N, but will be given by

$$N' = 1 + \frac{e_r^2}{4kT'RB} = 1 + (N-1)\frac{T_0}{T'}, \qquad \qquad \ldots \ldots (15)$$

which is greater or less than N according as T' is less or greater than T_0. In figure 2, the effective noise factor N' is shown as a function of T'/T_0 for various values of N. It is readily seen that the values of T'/T_0 which occur in practice result in a considerable improvement in the effective noise factor, e.g. if for a given receiver $N = 25$ (14 db.), then for $T'/T_0 = 10$ it is found that $N' = 5.3$ db.,

---

* Since the present paper was written, Reber (1944) and Southworth (1945) report measurements of solar noise, using highly directional aerials which give intensities of the same order as the black-body value calculated above.

while for $T''/T_0 = 100$, $N' = 0.9$ db. This serves to explain why in practice an improved design of receiver or aerial frequently fails to give the expected improvement in sensitivity.
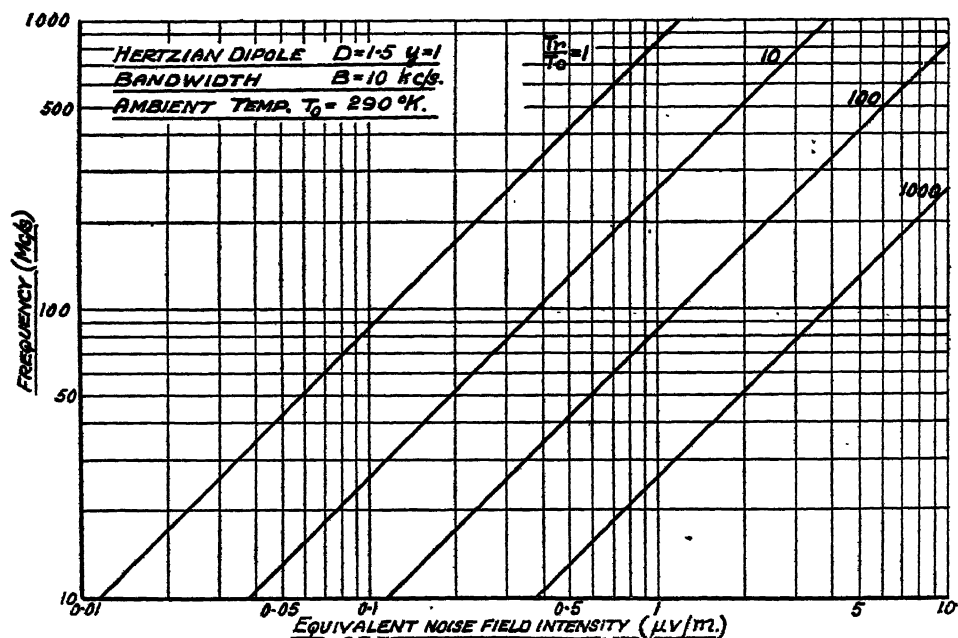
Figure 1. Equivalent noise field intensity for a dipole receiving radiation of temperature T_r.
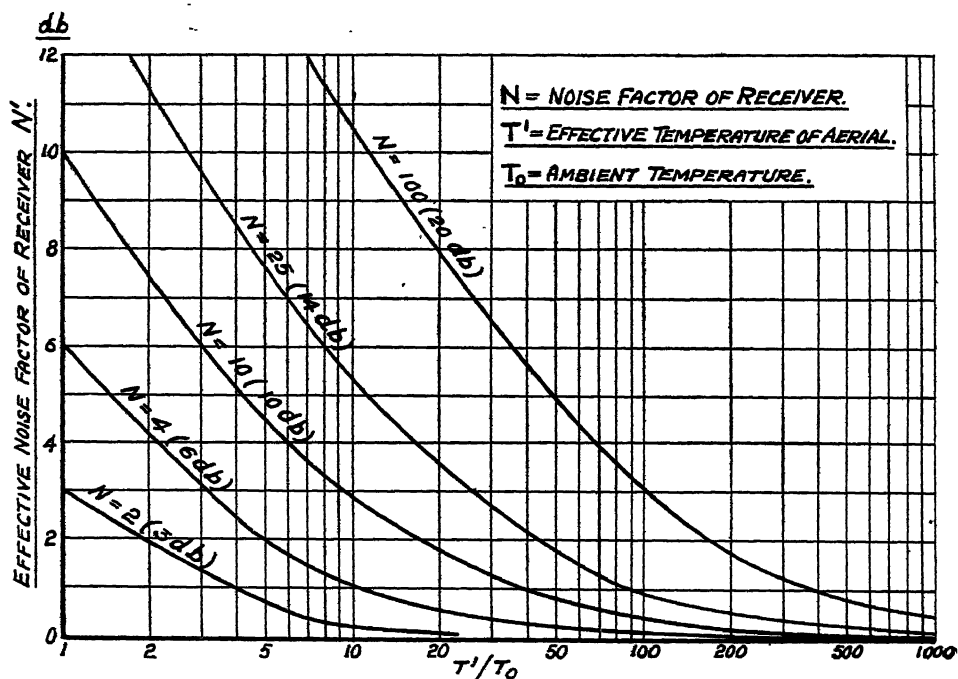
Figure 2. The influence of the aerial noise on the effective noise factor of a receiver.

The effect of receiver noise on the observation of aerial noise can readily be estimated. A convenient method of determining the aerial noise (i.e. T'') is to

measure the ratio n of the output noise powers when the receiver is connected in turn to the aerial (at T') and to the dummy (at T_0):

$$n = \frac{4kT'RB + e_r{}^2}{4kT_0RB + e_r{}^2} = 1 + \frac{T' - T_0}{NT_0}. \qquad \qquad \dots\dots(16)$$

Figure 3 shows the noise ratio n as a function of T'/T_0 for various values of N. The accuracy of the measurement of T' can be expressed by means of the differential ratio:

$$\frac{\Delta T'}{T'} \Big/ \frac{\Delta n}{n} = 1 + (N - 1)\frac{T_0}{T'} = N'. \qquad \qquad \dots\dots(17)$$

Thus the fractional uncertainity of T' is always greater than that in n, and if $\Delta T'/T'$ becomes the order of unity it may be said that the aerial noise is not measurable.
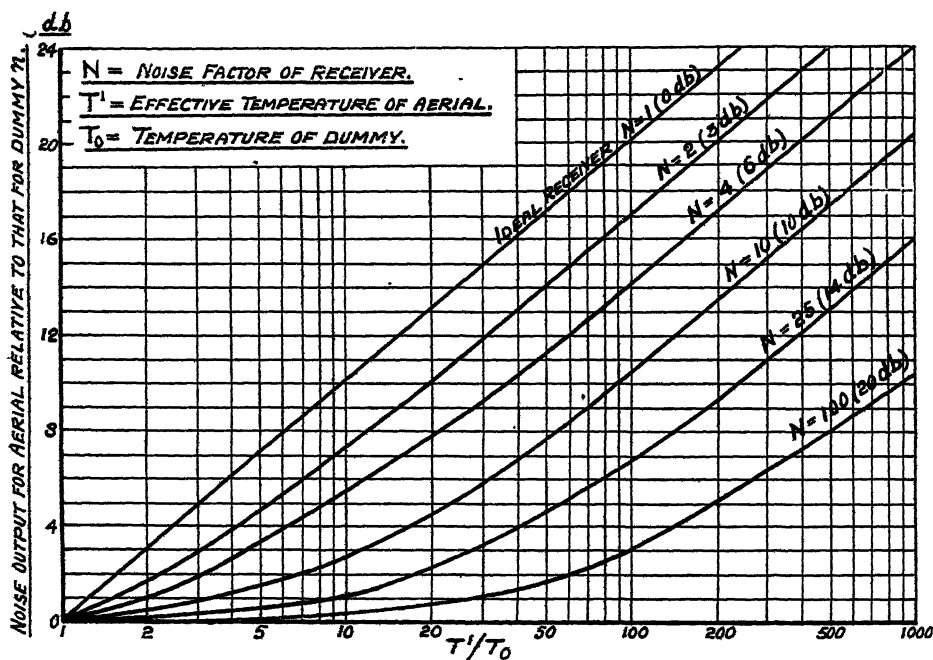
Figure 3. The influence of receiver noise on the measurement of aerial noise.

As the frequency of measurement increases, it is possible to obtain higher aerial directivity for a given aperture and thus the effective aerial temperature T' due to noise from a single source (such as the sun) is raised. This improvement is, however, slightly offset by the increase in the receiver noise factor at the higher frequencies.

<div align="center">ACKNOWLEDGMENT</div>

REFERENCES

BURGESS, R. E., 1941. " Noise in Receiving Aerial Systems." *Proc. Phys. Soc.*, **53**, 293.

FRANZ, K., 1942. " The Measurement of Receiver Sensitivity on Short Waves." *Hochfrequenztech. Elekt.*, **59**, 143.

HENYEY, L. G. and KEENAN, P. C., 1940. " Interstellar Radiation from Free Electrons and Hydrogen Atoms." *Astrophys. J.*, **91**, 625.

JANSKY, K. G., 1935. " A Note on the Source of Interstellar Interference." *Proc. Inst. Radio Engrs.*, **23**, 1158.

JANSKY, K. G., 1937. " Minimum Noise Levels obtained on Short-wave Radio Receiving Systems." *Proc. Inst. Radio Engrs.*, **25**, 1517.

REBER, G., 1940 a. " Cosmic Static." *Proc. Inst. Radio Engrs.*, **28**, 68.

REBER, G., 1940 b. " Cosmic Static." *Astrophys. J.*, **91**, 621.

REBER, G., 1942. " Cosmic Static." *Proc. Inst. Radio Engrs.*, **30**, 267.

REBER, G., 1944. " Cosmic Static." *Astrophys. J.*, **100**, 279.

SOUTHWORTH, G. C., 1945. ." Microwave Radiation from the Sun." *J. Franklin Inst.*, **239**, 285.

MEASUREMENT OF DIFFUSION AND THERMAL CONDUCTIVITY " CONSTANTS " IN NON-HOMOGENEOUS MEDIA, AND IN MEDIA WHERE THESE " CONSTANTS " DEPEND RESPECTIVELY ON CONCENTRATION OR TEMPERATURE

BY R. M. BARRER,
Technical College, Bradford

ABSTRACT. Equations have been derived for the steady-state distributions of temperature or concentration across membranes in the form of plates, tubes or spherical shells, when the thermal conductivity, or diffusion, " constants " are functions respectively of temperature, or solute concentration, in the diffusion medium, and also where these " constants " are functions of the positional co-ordinates. Equations are similarly given for the quantity of heat or solute transported through the membranes.

Procedures based on these equations are given for measuring diffusion and thermal conductivity " constants " and their dependence upon concentration and temperature respectively, or on positional co-ordinates. All observational data are based only on the steady state of flow, thus avoiding the need to deal with equations such as $\dfrac{\partial c}{\partial t} = \dfrac{\partial}{\partial x}\left(D\dfrac{\partial c}{\partial x}\right)$, which have so far not been integrated. The methods are suitable for measurements of flow of liquids, vapours or dyestuffs through natural and artificial plastics, of gases and liquids through zeolitic media, in surface migration of adatoms on metals, and similar problems.

§ 1. INTRODUCTION

THE focus of attention in quantitative studies of the flow of matter or of heat has long been Fick's law :

$$\frac{\partial c}{\partial t} = D\nabla^2 c, \qquad \qquad \ldots\ldots(1)$$

$$\frac{\partial T}{\partial t} = D\nabla^2 T, \qquad \qquad \ldots\ldots(1\,a)$$

where c and T are concentration and temperature respectively, and D is the diffusion or thermal diffusivity constant. It is now clear that D is often not a

constant at all, so that this law is a limiting law only, just as in other spheres are the Langmuir isotherm, the Debye-Hückel equation or Raoult's law.   Equation (1) must then be replaced by the following equation, which allows for the non-constancy of D:

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D\,\frac{\partial c}{\partial x}\right) + \frac{\partial}{\partial y}\left(D\,\frac{\partial c}{\partial y}\right) + \frac{\partial}{\partial z}\left(D\,\frac{\partial c}{\partial z}\right). \qquad \ldots\ldots(2)$$

No exact solution of equation (2) has been found for any boundary conditions, although in several instances approximate solutions have been derived (e.g. Hopkins, 1938).   Variants of a graphical method have been evolved (Matano, 1933; Eversole, Peterson and Kindsvater, 1938) for determining D as a function of c from the equation

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D\,\frac{\partial c}{\partial x}\right). \qquad \ldots\ldots(2\,a)$$

The method has been used by Wells and Mehl (1941), Matano (1933), Mehl (1937) and Rhines and Mehl (1937) for diffusion in metallic media and for diffusion of salts in solutions by Eversole *et al.*, 1941, 1942).

In general D may be a function of concentration of the diffusing species, or a function of the positional co-ordinates, x, y and z. D depends on c when metals and non-metals diffuse into metals (see above references; also Wells and Mehl, 1940 ; Smoluchowski, 1942, 1943 ; Barrer, 1941 a);   salts diffuse in solution (Eversole *et al.*, 1941, 1942); gases and liquids diffuse in zeolitic media (Barrer, 1941 b; Barrer and Ibbitson, 1944); adatoms diffuse on metallic surfaces (Langmuir, 1932, 1934; Langmuir and Kingdon, 1925; Bosworth, 1935, 1937); and liquids or vapours are imbibed by polymers which thereby swell (Daynes, 1937; Barrer, 1941 c; Garvie and Neal, 1938; King, 1945). D depends on positional co-ordinates when heat or matter flows in physically or chemically inhomogeneous media.   The unequal thermal gradients which arise at different places on the earth's crust (Bullard, 1945), or the permeability of inhomogeneous rock strata to fluids such as petroleum, afford examples of interest.   Diffusion of thorium along a tungsten filament where a temperature gradient exists along the filament is a more complex example, for here D depends both on the thorium concentration (Langmuir, 1932, 1934) and also on the positional co-ordinate, x.   Dependence on x in such problems arises independently of dependence on c because D is a function of $T\left(D=D_0 e^{-\frac{E}{RT}}\right)$ and T depends on x.

The rather laborious character of the graphical method of measuring D as a function of c, and the absence of a procedure when $D=f(x,y,z)$, render it important to evolve new simple procedures.   The theory of diffusion in condensed phases often awaits methods which permit data to be readily obtained.   In this paper procedures are developed for measuring D when this " constant " depends either on concentration or on the positional co-ordinates.   The new methods require observation of the stationary state of flow only, with a corresponding increase in directness and practicality.   They apply equally to the diffusion of heat or of matter.

§2. THE FUNDAMENTAL EQUATIONS

In Cartesian co-ordinates, equation (2) is the fundamental expression for material diffusion; its counterpart for flow of heat is

$$C_p\rho \cdot \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(K\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(K\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(K\frac{\partial T}{\partial z}\right). \qquad \ldots\ldots(2\,b)$$

In these equations D has the dimensions $l^2 t^{-1}$, and K, the *thermal conductivity constant*, the dimensions $m^1 l^{-1} t^{-1}$. Also C_p is the heat capacity at the point (x, y, z), and ρ the density at this point. Equation (2) is thus a particular case of $(2\,b)$, where $C_p\rho$ is independent of the dependent or independent variables. However, when D and K are true constants, equations (2) and $(2\,b)$ reduce to equations (1) and $(1\,a)$ respectively, the *thermal diffusivity constant* $D = K/C_p\rho$ in equation $(1\,a)$ now having the same dimensions as the *diffusion constant* D in equation (1).

Transformation of equations (2) and $(2\,b)$ to cylindrical co-ordinates gave for flow of heat and matter, respectively,

$$C_p\rho \cdot \frac{\partial T}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(rK\frac{\partial T}{\partial r}\right) + \frac{1}{r^2}\frac{\partial}{\partial \theta}\left(K\frac{\partial T}{\partial \theta}\right) + \frac{\partial}{\partial z}\left(K\frac{\partial T}{\partial z}\right), \qquad \ldots\ldots(3)$$

$$\frac{\partial c}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(rD\frac{\partial c}{\partial r}\right) + \frac{1}{r^2}\frac{\partial}{\partial \theta}\left(D\frac{\partial c}{\partial \theta}\right) + \frac{\partial}{\partial z}\left(D\frac{\partial c}{\partial z}\right), \qquad \ldots\ldots(3\,a)$$

Transformation to polar co-ordinates gave similarly

$$C_p\rho \cdot \frac{\partial T}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 K\frac{\partial T}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta K\frac{\partial T}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial \phi}\left(K\frac{\partial T}{\partial \phi}\right),$$
$$\ldots\ldots(4)$$

$$\frac{\partial c}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D\frac{\partial c}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta D\frac{\partial c}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial \phi}\left(D\frac{\partial c}{\partial \phi}\right).$$
$$\ldots\ldots(4\,a)$$

Many important problems concern diffusion in cylinders and spheres with concentration or temperature independent of θ and z (for cylinders) or of θ and ϕ (for spheres), so that the corresponding equations follow from equations (3) and (4) on omitting terms in these variables. In the steady state in which we are interested $\frac{dc}{dt} = \frac{dT}{dt} = 0$, and for heat flow, for example, one has

$$0 = \frac{d}{dr}\left(r^2 K\frac{dT}{dr}\right) \quad \text{(sphere)}, \qquad \ldots\ldots(5)$$

$$0 = \frac{d}{dr}\left(rK\frac{dT}{dr}\right) \quad \text{(cylinder)}. \qquad \ldots\ldots(6)$$

To these may be added the equation for heat flow in the x-direction only:

$$0 = \frac{d}{dr}\left(K\frac{dT}{dr}\right) \quad \text{(plate)}, \qquad \ldots\ldots(7)$$

The corresponding equations for diffusion of matter follow at once by substituting D for K and c for T.

The equations (5), (6) and (7) may be integrated readily. Thus for equation (5)

$$r^2 D \frac{dc}{dr}, \quad \text{or} \quad r^2 K \frac{dT}{dr} = A, \qquad \dots\dots(8)$$

where A is a constant. Now D (or K) may be a function of c or of $r(x)$. In the former case one has quite generally

$$D = D_0 [1 + f(c)] \qquad \dots\dots(9)$$

and in the latter

$$D = D_0 [1 + f(r)]. \qquad \dots\dots(10)$$

In equation (9), $f(c)$ is conventionally chosen so that D_0 is the value of D when $c \to 0$; in equation (10), similarly, D_0 is the value of D at a chosen face $r = r_1$. In either case the variables may be separated, and one obtains for the steady-state distribution of *solute* across a spherical shell:

$$D_0 \int [1 + f(c)] dc = -\frac{A}{r} + B \qquad \dots\dots(11)$$

when D is a function of c, and

$$D_0 c = \int \frac{dr}{r^2 [1 + f(r)]} + B \qquad \dots\dots(12)$$

when D is a function of r. Similarly for the distribution across a hollow cylindrical tube

$$D_0 \int (1 + f(c)) dc = A \ln r + B, \qquad \dots\dots(13)$$

$$D_0 c = A \int \frac{dr}{r [1 + f(r)]} + B, \qquad \dots\dots(14)$$

while for a plate

$$D_0 \int [1 + f(c)] dc = Ax + B \qquad \dots\dots(15)$$

$$D_0 c = A \int \frac{dx}{[1 + f(x)]} + B. \qquad \dots\dots(16)$$

§3. DETERMINATION OF D_0 AND $f(c)$ WHEN $D = D_0[1 + f(c)]$

Equations (11), (13) and (15) all provide means of measuring both D_0 and $f(c)$ when $D = D_0[1 + f(c)]$.

When the boundary conditions for diffusion across the plate are

$$c = c_1 \text{ for all } t \text{ at } x = 0,$$
$$c = c_2 \text{ for all } t \text{ at } x = 1,$$

equation (15) for the stationary state becomes

$$\frac{c_1 + F(c_1) - (c + F(c))}{c_1 + F(c_1) - (c_2 + F(c_2))} = \frac{x}{l}, \qquad \dots\dots(17)$$

where* $F(c) = \int f(c) \, dc$. Similarly for the tube or spherical shell with boundary conditions

$$c = c_1 \text{ for all } t \text{ at } r = r_1,$$
$$c = c_2 \text{ for all } t \text{ at } r = r_2,$$

---

* Integration is between the limits 0 and c. $F(c)$ thus tends to zero as $c \to 0$, the laws of flow in the limit then reducing to the simple Fick laws in accord with experiment.

one obtains from equation (13) (the tube)

$$\frac{c_1 + F(c_1) - (c + F(c))}{c_1 + F(c_1) - (c_2 + F(c_2))} = \frac{\ln r_1 - \ln r}{\ln r_1 - \ln r_2} \qquad \ldots\ldots(18)$$

and from equation (11) (the spherical shell)

$$\frac{c_1 + F(c_1) - (c + F(c))}{c_1 + F(c_1) - (c_2 + F(c_2))} = \frac{r_2}{r_1 - r_2} \cdot \frac{r_1 - r}{r}. \qquad \ldots\ldots(19)$$

Thus the steady state concentration distributions depend on the way in which D depends on concentration. From this distribution $F(c)$ and so $f(c) = \frac{d}{dc}(F(c))$ may be derived. Alternatively, one may derive $\frac{dc}{dx}$, which, when $c_2 = 0$, is (from equation (17) for example) $-\frac{dc}{dx} = \frac{c_1 + F(c_1)}{l(1 + f(c))}$. Then on plotting $-\frac{dc}{dx}$ as a function of c, one sees the intercept on the axis of x to be $\frac{c_1 + F(c_1)}{l}$, and from values of $\frac{dc}{dx}$ between c_1 and $0, f(c)$ easily follows.

D_0 in equation (9) can now be obtained by measuring the steady-state flow of solute through the medium. This flow per unit area and per unit time is

$$P = -\left(D\frac{dc}{dx}\right)_{x=l}. \qquad \ldots\ldots(20)$$

By differentiation of equation (17) with respect to x one finds $\frac{dc}{dx}$, and then by integration of $\left(D\frac{dc}{dx}\right)_{x=l}$ from 0 to t the quantity Q of solute passing in time t through a membrane of area A is found:

$$Q = A\int_0^t P dt = \frac{D_0}{l}\left[c_1 + F(c_1) - (c_2 + F(c_2))\right]At. \qquad \ldots\ldots(21)$$

For a tube of length L a similar procedure gives

$$Q = 2\pi r_2 L\int_0^t P dt = 2\pi L D_0\left[\frac{c_1 + F(c_1) - (c_2 + F(c_2))}{\ln r_1 - \ln r_2}\right]t, \qquad \ldots\ldots(22)$$

and for a spherical shell

$$Q = 4\pi r_2^2\int_0^t P dt = 4\pi\frac{r_1 r_2}{r_1 - r_2} D_0[c_1 + F(c_1) - (c_2 + F(c_2))]t. \qquad \ldots\ldots(23)$$

Since $F(c)$ has previously been determined using equations (17), (18), or (19), D_0 can be evaluated in the appropriate equation (21), (22), or (23).

An alternative procedure for determining D_0 and $f(c)$ makes use only of equations (21), (22) or (23). When $c_2 \ll c_1$, equation (21), for instance, becomes

$$\frac{Ql}{tA} = D_0[c_1 + F(c_1)], \qquad \ldots\ldots(21\,a)$$

where all quantities on the left-hand side of equation (24 a) are measured.   If, now, c_1 is made small, departures from the simple Fick laws tend to zero, so that

$$\lim_{c_1 \to 0} \frac{Ql}{tAc_1} = D_0. \qquad \ldots\ldots (21\,b)$$

Next $F(c)$ is given by the relation

$$\frac{Ql}{tAD_0} - c_1 = F(c_1), \qquad \ldots\ldots (24)$$

the left-hand side of which is determined for a range of values of c.   Finally

$$f(c) = \frac{d}{dc}(F(c)).$$

§4. DETERMINATION OF D_0 AND $f(x)$ WHEN $D=D_0(1+f(x))$

We now return to consideration of equations (12), (14) and (16) for the steady-state distribution of c when $D = D_0[1 + f(x)]$.   If the integral on the right-hand side of each of these equations be denoted by I, and by I_1 and I_2 at each boundary surface, the constants A and B may be eliminated to give for the boundary conditions in the previous section

$$\frac{c_1 - c}{c_1 - c_2} = \frac{I_1 - I}{I_1 - I_2}. \qquad \ldots\ldots (25)$$

Moreover, the quantity Q of solute diffusing through the outgoing surface of a plate of area A is

$$Q = A\int_0^t -\left(D\frac{dc}{dx}\right)_{x=l} dt = -AD_0 t.\frac{c_1 - c_2}{I_1 - I_2}, \qquad \ldots\ldots (26)$$

while for a hollow tube of length L

$$Q = 2\pi r_2 L\int_0^t -\left(D\frac{dc}{dr}\right)_{r=r_2} dt = -2\pi L.\frac{c_1 - c_2}{I_1 - I_2} \qquad \ldots\ldots (27)$$

and for the spherical shell

$$Q = 4\pi r_2^2 \int_0^t -\left(D\frac{dc}{dr}\right)_{r=r_2} dt = -4\pi D_0 t.\frac{c_1 - c_2}{I_1 - I_2}. \qquad \ldots\ldots (28)$$

Then to determine D_0 and $f(x)$ for the plate, for example, the steady-state concentration distribution can be measured, and from it the value of

$$\frac{dc}{dx} = \frac{c_1 - c_2}{I_1 - I_2}\frac{1}{(1 + f(x))}.$$

On plotting $\dfrac{dc}{dx}$ against x, its value at $x=0$ is $\dfrac{c_1 - c_2}{I_1 - I_2}$, and from its values for $0 < x < 1$, $f(x)$ may be found.   Equation (26) then serves to determine D_0.

Details given in this and the previous section apply equally well to the determination of K_0 and $f(T)$ or $f(x)$ in the analogous problem of heat flow.

§ 5. CHARACTER OF STEADY-STATE CONCENTRATION DISTRIBUTION WHEN D DEPENDS ON c OR ON x

In equations (9) or (10) the functions $f(c)$ and $f(x)$ (or $f(r)$) must be such that D is never negative or infinite. For a suitable choice of constants appropriate functions include

$$f(c) = \pm ac^n \pm bc^m \pm dc^p \ldots \text{(exponents positive or zero)}, \quad \ldots\ldots(29)$$

$$f(c) = \pm \frac{ac}{1+bc} \quad \ldots\ldots(30)$$

$$f(c) = \pm ae^{\pm bc}, \quad \ldots\ldots(31)$$

$$f(c) = \pm ac^n e^{\pm b/c} \quad (n \geqslant 0). \quad \ldots\ldots(32)$$

Atom % Sn, Be, Si or Zn
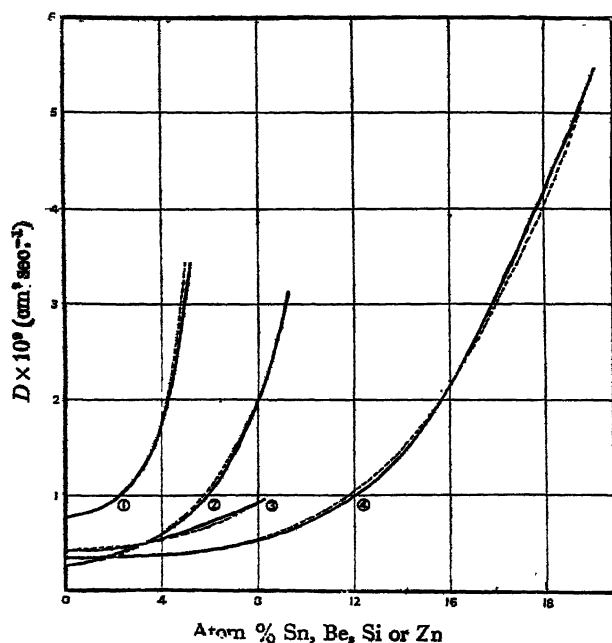
Figure 1 a. Experimental (Rhines' and Mehl, 1938) (full curve) and calculated (dotted curve) relations between D and c. Calculated curves employ the exponential functions $D = D_0(1 + ae^{bc})$ and $D = D_0(1 + ac^2 e^{-b/c})$.

Calculated Curve (1): Sn—Cu. $D = 7 \cdot 7 \times 10^{-10}(1 + 4 \cdot 46 \times 10^{-2} e^{0 \cdot 862c})$ cm^2 sec.$^{-1}$.

,,   Curve (2): Si — Cu. $D = 2 \cdot 31 \times 10^{-10}(1 + 0 \cdot 306 \, e^{0 \cdot 402c})$ cm^2 sec.$^{-1}$.

,,   Curve (3): Be—Cu. $D = 4 \cdot 23 \times 10^{-10}(1 + 7 \cdot 08 \times 10^{-2} e^{0 \cdot 852c})$ cm^2 sec.$^{-1}$.

,,   Curve (4): Zn—Cu. $D = 3 \cdot 46 \times 10^{-10}\left(1 + 1 \cdot 51 \times 10^{-3} c^3 e^{-\frac{2 \cdot 304}{c}}\right)$ cm^2 sec.$^{-1}$.

Figures 1 a and 1 b give actual and calculated D-c curves according to several of these equations, and illustrate the degree of success obtained with them.

The steady-state concentration distribution* across a plate of the diffusion medium is given by equation (20), which for $c_2 \sim 0$ becomes

$$c + F(c) = \left(1 - \frac{x}{l}\right)[c_1 + F(c_1)]. \quad \ldots\ldots(17a)$$

* In the absence of rate-controlling phase-boundary processes (For a discussion of these processes see Barrer, 1941 a).

It is convenient to choose units for c such that $c_1 = 1$; then in equations (29) to (32), $0 < c < 1$. Also, if the solute has a limited solubility in the diffusion medium, c in these equations may conveniently be replaced by c/c_s where c_s denotes the saturation concentration. Using these conventions, figures $2a$ and $2b$ give typical concentration distributions for some of the functions of equations (29) to (32). When D increases as c increases, equilibrium distributions are convex away from the axis of x; but when D decreases with increasing c, the distributions are concave towards that axis. When D at first increases and then decreases, or vice versa,
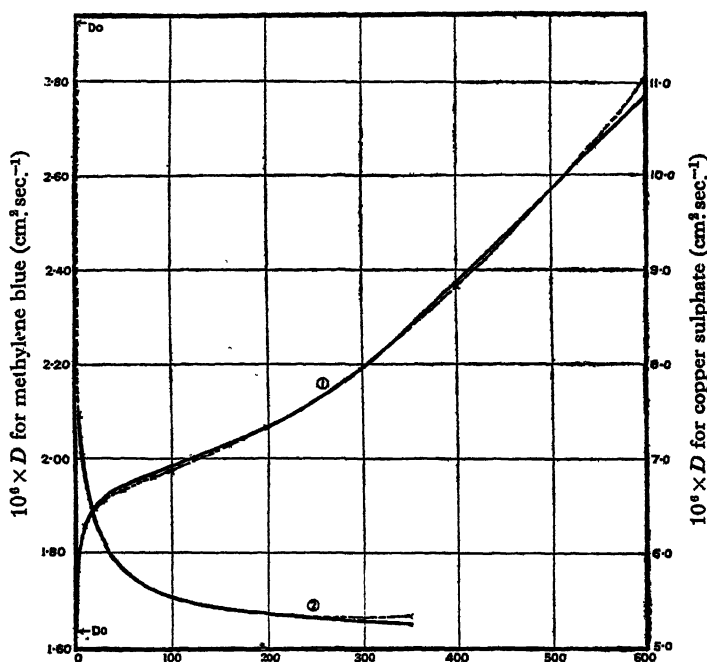
Concentration (millemoles/litre for $CuSO_4$; mgrm./litre for methylene blue.)

Figure 1 *b*. Experimental (Eversole *et al.*, 1941, 1942), (full curve) and calculated (dotted curve) relations between D and c. Calculated curves employ the functions $D = D_0(1 + ac^n + bc^m + \ldots)$.

Calculated curve 1 : Methylene blue—H_2O.
$$D = 1 \cdot 635 \times 10^{-6}(1 + 0 \cdot 0824c^{1/4} - 0 \cdot 00128c + 0 \cdot 0000738c^{3/2}) \quad cm^2 \, sec.^{-1}.$$

Calculated curve 2 : Copper sulphate—H_2O.
$$D = 11 \cdot 65 \times 10^{-6}(1 - 0 \cdot 00311c^{1/4} + 0 \cdot 002645c^{1/3}) \quad cm^2 \, sec.^{-1}.$$

with increasing c, the behaviour is less predictable (e.g. figure $2b$, curve 3). Curves 4 and 6 in figure $2a$ give the extreme range of steady-state distributions of concentration for a pure zeolitic diffusion process where (*cf.* equation (29)) $D = D_0(1 - c/c_s)$.

Equilibrium distributions when $D = D_0[1 + f(x)]$ follow from equation (25). Two typical values for $f(x)$ will be considered:

$$f(x) = ax, \qquad \qquad \ldots\ldots (33)$$

$$f(x) = bx + ax^2. \qquad \qquad \ldots\ldots (34)$$

When $f(x) = ax$, I in equation (25) has the values

$$I = \frac{1}{a}\ln(1+ax) \quad \text{(for plate)}, \qquad \ldots\ldots(35)$$

$$I = -\ln\frac{1+ar}{r} \quad \text{(for tube)}, \qquad \ldots\ldots(36)$$

$$I = -\frac{(1+ar)}{r} + a\ln\frac{(1+ar)}{r} \quad \text{(for spherical shell)}, \qquad \ldots\ldots(37)$$

2 a.

2 b.
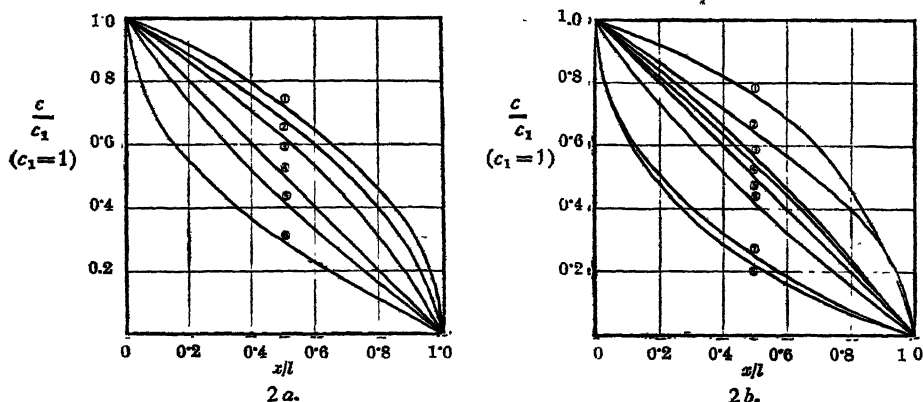
Figure 2 a. Typical steady-state concentration distributions across a membrane when $D = D_0(1+f(c))$.

Curve (1): $f(c) = ac$; $a = 100$.
Curve (2): $f(c) = ac$; $a = 10$.
Curve (3): $f(c) = ac$; $a = 2$.
Curve (4): $f(c) = 0$ (simple Fick law obeyed).
Curve (5): $f(c) = -ac$; $a = 0.5$.
Curve (6): $f(c) = -ac$; $a = 1.0$.

Figure 2 b. Typical steady-state concentration distributions across a membrane when $D = D_0(1+f(c))$.

Curve (1): $f(c) = ae^{bc}$; $a = 1$, $b = 3$.
Curve (2): $f(c) = \dfrac{ac}{1+bc}$; $a = 100$, $b = 1$.
Curve (3): $f(c) = -ac^{1/2} + bc^2$; $a = 1$, $b = 2$.
Curve (4): $f(c) = \dfrac{ac}{1+bc}$; $a = 1$, $b = 1$.
Curve (5): $f(c) = 0$ (simple Fick law obeyed).
Curve (6): $f(c) = -\dfrac{ac}{1+bc}$; $a = 0.9$, $b = 1$.
Curve (7): $f(c) = -ac^{1/2}$; $a = 1$.
Curve (8): $f(c) = -ac^{1/4}$; $a = 1$.

and values for I_1 and I_2 follow on setting $x = 0$ and $x = 1$ for the plate, or $r = r_1$ and r_2 for the tube and spherical shell. When $f(x) = bx + ax^2$, several values of

$$I = \int \frac{dx}{1 + bx + ax^2}$$

result for different values of a and b:

$$I = \frac{2}{\sqrt{4a - b^2}} \tan^{-1} \frac{2ax + b}{\sqrt{4a - b^2}} \ (4a > b^2),$$

$$I = \frac{1}{\sqrt{b^2 - 4a}} \ln \frac{2ax + b - \sqrt{b^2 - 4a}}{2ax + b + \sqrt{b^2 - 4a}} \ (b^2 > 4a),$$

$$I = -\frac{2}{2ax + b} \ (b^2 = 4a),$$

$\left.\right\}$ (for plate) ......(38),
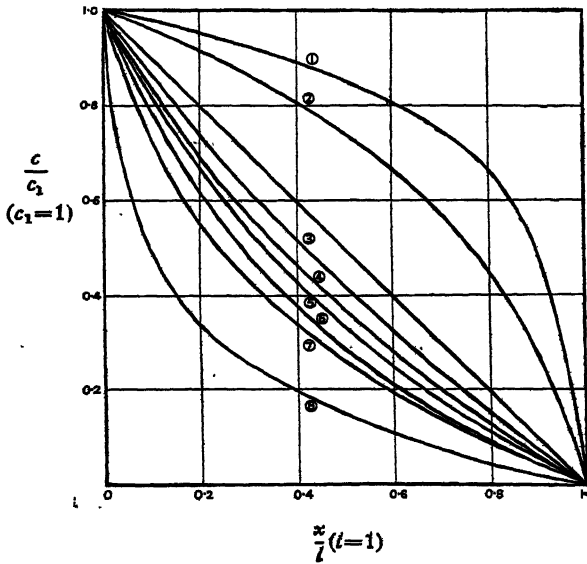
Figure 3. Typical steady-state concentration distributions when $D = D_0(1 + f(x))$. The straight line corresponds to thes simple Fick law ($f(x) = 0$).

Curve (1): $f(x) = -ax$; $a = 0.99$.
Curve (2): $f(x) = -ax$; $a = 0.90$.
Curve (3): $f(x) = ax$; $a = 1.0$.
Curve (4): $f(x) = ax$; $a = 2.0$.
Curve (5): $f(x) = bx + ax^2$; $a = 1, b = 2$.
Curve (6): $f(x) = bx + ax^2$; $a = 2.25, b = 3$.
Curve (7): $f(x) = ax$; $a = 9$.
Curve (8): $f(x) = ax$; $a = 99$.

and I_1 and I_2 follow as before. Similarly, when $f(r) = br + ar^2$, the value of

$$I = \int \frac{dr}{r(1 + br + ar^2)}$$

for the tube is

$$I = \tfrac{1}{2} \ln \left(\frac{r^2}{1 + br + ar^2} \right) - \frac{b}{2} I_{\text{plate}}, \qquad \ldots\ldots(39)$$

where I_{plate} denotes the appropriate integral for the plate but with r replacing x. Finally, for the spherical shell,

$$I = \int \frac{dr}{r^2(1 + br + ar^2)}$$

is $$I = \frac{b}{2} \ln \frac{1 + br + ar^2}{r^2} - \frac{1}{r} + \frac{b^2 - 2a}{2} I_{plate}. \quad \quad \ldots \ldots (40)$$

Figure 3 shows some steady state concentration distributions across a plate of unit thickness when $c_1 = 1$ and $c_2 = 0$; and $D = D_0(1 + ax)$ (curves 8, and 1 to 5) and $D = D_0(1 + bx + ax^2)$ (curves 6 and 7). When D is an increasing function of x the curves are concave towards the axis of x, and when D is a decreasing function of x the $c : x$ curves are convex away from this axis.*

There is little material available for quantitative interpretation according to methods developed in this paper. In rubber, steady-state concentration distributions of water concave towards the x-axis (Taylor, Herman and Kemp, 1936), suggest that D decreases as the moisture content grows.† There seems little regularity in available data, however, for in other polymers, such as viscose silk and wool, D for certain diffusing species increases with c (Garvie and Neal, 1938; King, 1945).

REFERENCES

BARRER, R. M., 1941 a. *Diffusion in and through Solids* (Camb. Univ. Press), chap. 5, and expecially references given there to the work of Bramley and co-workeis.

BARRER, R. M., 1941 b. *Trans. Faraday Soc.*, 37, 590.

BARRER, R. M., 1941 c. *Diffusion in and through Solids* (Camb Univ Press) pp. 443–7.

BARRER, R. M. and IBBITSON, D., 1944. *Trans. Faraday Soc.*, 40, 206.

BOSWORTH, R. C., 1935. *Proc. Roy. Soc.*, 150 A, 58 ; 1936. *Ibid.*, 154 A, 112 ; 1937. *Ibid.*, 162 A, 32.

BULLARD, E. C., 1945. *Nature, Lond.*, 156, 35.

DAYNES, H., 1937. *Trans. Faraday Soc.*, 33, 531.

EVERSOLE, W., PETERSON, J. and KINDSVATER, H., 1941. *J. Phys. Chem.*, 45, 1398.

EVERSOLE, W., KINDSVATER, H. and PETERSON, J., 1942. *J. Phys. Chem.*, 45, 370.

GARVIE, W., and NEAL, S., 1938. *Trans. Faraday Soc.*, 34, 335.

HOPKINS, M., 1938. *Proc. Phys. Soc.*, 50, 703.

KING, G., 1945. *Trans. Faraday Soc.*, 41, 325.

LANGMUIR, I. and KINGDON, K., 1925. *Proc. Roy. Soc.*, 107 A, 61.

LANGMUIR, I., 1932. *J. Amer. Chem Soc.*, 54, 1252.

LANGMUIR, I., 1934, *Acta Physicochim., U.S.S.R.*, 1, 371

MATANO, C., 1933. *Jap. J. Phys.*, 8, 109 ; *Proc. Math. Soc. Jap.*, 15, 405.

MEHL, R., 1937. *J. App. Phys.*, 8, 174.

RHINES, F. and MEHL, R., 1937. *Trans. Amer. Inst. Min. Met. Engrs.*, 124, 331.

RHINES, F. and MEHL, R., 1938. *Trans. Amer. Inst. Min. Met. Engrs*, 128, 185.

SMOLUCHOWSKI, R., 1942. *Fhys. Rev.*, 62, 539 ; 1943. *Ibid.*, 63, 438.

TAYLOR, R., HERRMANN, D. and KEMP, A., 1936. *Ind. Eng. Chem.*, 28, 1255.

WELLS, C., and MEHL, R., 1940. *Metals Technol., A.I.M.E.*, Tech Pub. No. 1180.

WELLS, C. and MEHL, R., 1941. *Amer. Inst. Min. Eng.*, Tech. Pub. No. 12.

* Also for equal values of c_2, c_1, and l, whenever the equilibrium concentration distribution is convex away from the x-axis, the membrane is more permeable than predicted by Fick's law with $D = D_0$; and when this distribution is concave towards the x-axis the membrane is less permeable than predicted by Fick's law with $D = D_0$.

† In this instance an alternative explanation is possible. See e.g. Barrer (1941c).

OBITUARY NOTICES

JOHN EDMUND CALTHROP

BY the sudden death in June 1945 of J. E. Calthrop, Senior Lecturer in Physics, Queen Mary College has lost one of the most loyal members of its staff, and his fellow-teachers have lost a singularly gentle and unexacting friend.

Calthrop was born in Lincoln in July 1891. He took his London B.Sc. degree in 1912 as a student at the Borough Road Training College, and after two years spent in teaching science at the Lincoln Technical School he enlisted in the Honourable Artillery Company. He was severely wounded at Beaumont Hamel in November 1916, while serving as an infantryman ; after he had recovered from his wounds he was given a commission in the Royal Regiment of Artillery, and he saw active service again in Flanders with a Heavy Battery until the armistice. Calthrop's war service was of a pattern with his whole life : he was the last man one would expect to see sitting in a job at the Base, or collecting decorations on lines of communication.

After the war he entered at Pembroke College, Cambridge, taking a Natural Science Tripos, and in 1921 he joined the Physics staff of Queen Mary (then East London) College, under Professor C. H. Lees. For a period after his appointment he did research work in general physics and heat, publishing a number of papers in the *Philosophical Magazine* and in our *Proceedings*, but he soon decided that teaching was his real vocation, and for the last twenty years of his life he devoted himself almost entirely to the training of students and to doing all he could for their general welfare. He had a real affection for the young, and he never spared himself in their service. His interest in teaching is shown in a number of notes he published in the *School Science Review*, and in a highly original course of *Advanced Experiments in Practical Physics* which appeared in 1938 and was very favourably reviewed both here and abroad. His last publication, at the end of 1943, was a *German Physics Reader*.

His chief handicap as a teacher was an unconquerable diffidence of manner. He built up a fine library of modern physics books in English, French and German, which he studied assiduously (and, to mention his only vice, under-scored freely), he took pains to find the best way of treating a subject and he prepared admirable lectures—which he then, in class, delivered in a low tone and an almost apologetic manner, as though he were ashamed of their content. As a result, although the best students got a lot from him, the less discerning tended to under-value his teaching, and I fear that he was often conscious of this. His most effective teaching was done in the laboratory and in informal talks with the students, to whom he willingly sacrificed much of his spare time.

Calthrop was one of the most transparently honest and least self-seeking of men. Although of a naturally retiring disposition, and a bachelor, he was—as I have known since his earliest contacts with my own family—entirely at his ease with very young children, and outstandingly happy in his relations with them.

From 1939 to 1945, while Queen Mary College was evacuated to Cambridge, Calthrop resided in King's College. He enjoyed immensely the privilege of sharing in the corporate life of a great Cambridge College, and his obvious integrity, his eagerness to help and his evident delight in his surroundings brought him many new friends. There is much comfort in the thought that his last years, spent in such congenial surroundings, were among the happiest of his life.   H. R. ROBINSON.

JOHN JOB MANLEY

MANLEY was born on 13 July 1863 near Wellington, Somerset, and the family migrated later to Devon. At the age of 25 he became Daubeny Curator of the Laboratory of Magdalen College, Oxford, and after the first three years he taught practical physics and analytical chemistry to undergraduates reading for the Honours School of Natural Science. His first research work at this laboratory was carried out in collaboration with V. H. Veley

and resulted in a series of papers on the preparation of anhydrous nitric acid, and on the physical properties of aqueous solutions of this acid and of sulphuric acid. This work, which was completed in 1905, was recognized by the award of an M.A. degree, *honoris causa*, in 1903.

Manley was typical of the best school of experimentalists—those who took a pride in carrying out operations with their own hands. His opportunities for research were increased by the award of a Research Fellowship in 1917, the year in which he took his M.A. He gave up his teaching work in 1923 and his curatorship in 1929, in which year, at the age of 66, he left Oxford and equipped a laboratory in Bournemouth. He took an Oxford D.Sc. in 1931. He continued work at Bournemouth until 1938 as the holder of a Research Fellowship and then privately until his death on 24 March 1946, which occurred after only a fortnight's illness.

After his partnership with Veley had ceased, Manley devised a differential densimeter for the study of sea-water ; a duplicate of this apparatus, mounted on gimbals, was successfully used by N. P. Campbell on a voyage from London to Colombo. Later, whilst standardizing a set of weights, Manley encountered certain difficulties which led him to devise his enclosed or " protected " balance beam (1910), and ultimately to the revision of Landolt's work on the apparent change in weight during chemical reaction (1912). For this revision he chose the straightforward reaction between barium chloride and sodium sulphate, instead of the somewhat complicated ones selected by Landolt. He found that the change in weight did not exceed one part in 100 million. His interest in the subject of the precision balance and the study of analytical weights continued up to his last illness, and several papers were published in the *Proceedings of the Physical Society*, the last of them, on the advantages of rhodium plating, only last year (*Proc. Phys Soc.* **57**, 136 (1945)).

His experimental skill, and in particular his mastery of the art of glass blowing, was successfully applied to the study of extremely thin wires for platinum thermometers, gases at low pressures, and the interaction between mercury and helium. His method of purifying phosphoric acid by ozonized air, worked out in 1922, was used to good effect in 1929 to 1934 in an extension of Baker's work on the intensive drying of benzene.

J. J. Manley's chief recreations were those of a quiet, modest man, who liked the open air—cycling, climbing, gardening and photography. He was a devout churchman, with an assured faith in things spiritual. Many will remember him as an unassuming worker, an interesting speaker, with that self-confidence which comes from certainty of knowing his subject. He was twice married, and is survived by his second wife and by his son, C. H. Manley, the City Analyst of Leeds and Wakefield. One of his two grandsons is a civil engineer.

RAOUL FREDERIC SCHMID

R. F. SCHMID, Fellow of the Physical Society, died in Budapest on 2 October 1943. He was carried away by a disease of tragic suddenness at the age of 39. A man of science, he worked and taught all his life. As a true scientist he professed the freedom of thought and human rights in an age when this was dangerous and difficult, trusting a better period would come. This period, however, he did not live to see. With a loving heart he helped everybody, teaching also through the example of his own life.

At the age of 22 he had graduated at three Universities and possessed the diplomas of chemical and economic engineer and the degree of Ph.D. He was the winner of several scholarships and, as a Rockefeller Foundation Fellow, worked in Chicago in 1932. Originally a chemist, he saw in molecular spectra the firm experimental ground for studying the fundamental problems of chemistry. Working in this field, he added something in nearly every domain of molecular spectroscopy with an admirable fertility. He united wide knowledge and exceptional memory with experimental skill and invention.

Through the inscrutable ways of providence, his collaborator, L. Gerö, was destined to follow him in the short space of one year and a half. He died, at the age of 35, a victim of the war, at a moment when he could already expect to have got to the beginning of the eagerly expected new period. Their names are now closely connected by the scientific work they have done together.

To name only some milestones of this work : a minute investigation of the CO spectrum in order to get as complete an analysis as possible of the term system of this molecule, of such importance from the point of view of organic chemistry ; pioneer work in discovery and interpretation of perturbations and predissociations ; establishment of the dissociation schemes for CO, CN, N_2, NO and diatomic hydride and deuteride molecules ; an interpretation of the thermochemical data of organic chemistry and the emphasizing of the importance of excited atomic states from the point of view of chemistry.

Those who part from this world continue to live in their works and in the hearts of those who knew and loved them.   J. G. VALATIN.

REVIEWS OF BOOKS

Mathematical Theory of Plasticity, by W. PRAGER. (Brown University, Providence, Rhode Island, 1944.)

A considerable literature now exists dealing with various aspects of the mechanics of solids beyond the elastic range. Only in one or two instances, however, as far as the writer is aware, has any attempt been made to co-ordinate the available information in the form of treatises dealing with the subject as a whole, and none of these treatises appears to merit the description of " mathematical theory of plasticity ".

The present manuscript therefore meets a considerable need, giving as it does a summary of the present state of knowledge of mechanics of solids beyond the elastic range, together with a study of the mathematical theory of plasticity as applied to certain types of ideal materials.

In the opening two chapters the analysis of stress and strain is developed in terms of both Lagrangian and Eulerian co-ordinates, and some properties of the stress and strain tensors are discussed, particularly in relation to certain invariants of the stress and strain systems. Expressions for " work " are obtained in terms of the stress and strain tensors. These analyses are the basis of the consideration of stress-strain relations in subsequent chapters.

The third chapter considers the relations between the stress and strain tensors for an isotropic elastic body. For this purpose the stress and strain tensors are decomposed into functions of the mean stress or strain and the stress and strain deviators. The equations of equilibrium and compatibility are derived, and the relations between the invariants of stress and strain are considered. An expression for strain energy is obtained.

The nature of elastic breakdown is discussed, with consideration of the various theoretical criteria, and experimental evidence supporting them.

In Chapter IV the stress-strain relations for a visco-plastic material (defined as a viscous fluid having a definite yield point) are derived from consideration of the corresponding relations for a viscous fluid, by the use of the tensor of " over stress " expressing the extent by which the yield limit has been exceeded. The relations corresponding to several arbitrary conditions of yielding are stated.

Experimental methods of determining the coefficient of viscosity of a visco-plastic material are described. As a limiting case of the visco-plastic material, the stress-strain relations of a perfectly plastic material are derived, and the experimental evidence of Taylor and Quinney in this connection is discussed.

Chapter V deals with the effects of strain hardening on the stress-strain relations of a visco-plastic or plastic body, suitable modification being made of invariants of stress and strain concerned in the relations to account for this effect. Both isotropic and anisotropic strain hardening are considered. The latter form is assumed to be such that the three principal planes of strain constitute three equivalent planes of asymmetry, and that rotations interchanging principal axes of strain do not change the value of invariants.

In Chapter VI, the stress-strain relations of composite material are derived from the relationship obtained in previous chapters for the relatively simple materials. Materials composed of two or more of the following elements are considered—elastic solids, viscous fluids, visco-plastic materials and perfectly plastic materials. These materials include

those of Voigt and Maxwell. The constituent materials are either placed in series, with superposition of deformations and the same stress; or in parallel, with superposition of stresses, and the same deformation. The modification of the stress-strain relations for a composite visco-plastic and elastic material used by F. Odquvist for the analysis of creep· is stated.

Chapter VII deals with the subject of materials displaying a gradual transition from the elastic to the plastic state. ·Geometrical methods of representation of stress and strain. deviations are first evolved. Using these methods of analysis, and making the assumption that materials of this type obey a general relation of the type

$$dq = dp \left\{ \frac{1 - (q_1^2)^n}{C^2} \right\},$$

where q is stress and p is strain, it is shown that the equation relating the stress $(S_1 k)$· and strain $(\epsilon_1 k)$ tensors is of the type

$$A \frac{\partial}{\partial t} \frac{(S_1 k)}{A} = 2\mu \dot{\epsilon}_1 k, \quad \text{where} \quad A^{2n} = 1 - \frac{(S_{11})^{2n}}{C}.$$

C, μ and n are constants, t is time and S_{11} a second-order invariant. Modification of the stress and strain relations for materials of this type undergoing isotropic or non-isotropic strain hardening is discussed.

In Chapters VIII to XI the author shows the application of the theory of plasticity to the problems of pure shear, torsion of a cylindrical body, and plane strain with and. without axial symmetry. In each case general formulae are first developed, and are subsequently applied to the cases of the elastic, visco-plastic, St. Venant and Prandtl-Reuss materials, and also to materials with gradual transition from the elastic to the plastic state. Where precise mathamatical solutions of the problems presented are not found possible, reference is made to various experimental methods of obtaining the values of stress or strain functions, including the flow, electric field, and membrane analogies. Numerous references are given to the publications of other workers who have dealt with these particular problems.

In the concluding chapter two boundary-value problems involving the variation of the stresses and displacements in certain areas of incompressible bodies, including elastic material, Prandtl-Reuss material and material with gradual transition to the plastic state, are considered. Solutions are obtained by the employment of variational principles postulating minimum values for certain combined functions of stresses and displacements.

Two minor criticisms have occurred to the reviewer. Regarding the content of the manuscript, it seems a matter of some regret that in view of the practical importance of the subject of creep or time-dependent strain, the author has given only passing notice to it.

Regarding the actual arrangement of the text, the provision of a glossary of symbols used would have improved the readability of the manuscript, since these symbols are quite numerous and constitute a considerable tax on the reader's memory.

It is to be hoped that in a future manuscript the author may apply the theory developed. in this present work to further problems of plasticity, many of which, e.g. those involved in the engineering design of gas turbines, pipe flanges, and aerofoils, are now acquiring· very considerable practical importance.   A. E. JOHNSON.

Mathematical Theory of Optics; by R. K. LUNEBERG. Pp. xvii–401. (Brown. University, Providence, Rhode Island, 1944.) \$4.

These mimeographed notes are a record of a course of lectures given at Brown University in the summer of 1944. The subject-matter is dealt with in a comprehensive and. unbiased manner, and the book should prove of value to those looking for a thorough treatment of the theory underlying image formation in, and the design of, optical systems. It is likely to be of interest to those designers who make use of trigonometric ray tracing· and are seeking an alternative method. No mention is made of such tracing, the theory· being developed from a point of view which differs fundamentally from, say, that adopted. in the well known book by Conrady. Considerable use is made of the characteristic·

function method and the power and breadth of application of this method is amply demonstrated. The book may be regarded as a companion to the Cambridge Tracts by Synge and by Steward and the classic, *Optik*, by Professor Max Born.

In the first chapter the wave theory (based on Maxwell's equations) and the ray theory of optics are developed simultaneously. Transparent isotropic media are considered. The kernel of this theory is the equation of the wave fronts (the eiconal equation),

$$\left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2 + \left(\frac{\partial\psi}{\partial z}\right)^2 = n^2(x, y, z),$$

the light rays being the orthogonal trajectories of a set of wave fronts, $\psi = ct$. Differential equations satisfied by the light rays are found and Fermat's principle is introduced. It is shown that wave fronts may be constructed with the aid of light rays ; and, conversely, by means of Jacobi's theorem, a method is obtained for constructing the light rays when a solution to the above equation is known. Applications are made to (a) spherical waves in a homogeneous medium, (b) wave fronts in a medium of discontinuous optical properties, (c) periodic waves of small wave-length.

Hamilton's theory is presented in Chapter II. The three characteristic functions, the point-, mixed-, and angle-functions, are defined and their properties and physical significance are studied. Integral invariants and their application to ray congruences are discussed and an interesting geometrical interpretation is given of Poincaré's integral invariant. There are a few examples of the characteristic functions, but the important applications are reserved for Chapter III, in which some of the topics studied are the correction of optical instruments by aspheric surfaces, axially symmetric systems, spherical aberration and coma, and the problem of designing condenser systems.

A comprehensive treatment of first-order optics is given in Chapter IV. Considerable use is made of a matrix notation in the manipulation of the mapping equations. The problem of first-order optics is to investigate the properties of the four-parameter system, S, of rays which lie in the neighbourhood of a given ray, for example the principal ray. This is shown to be equivalent to a study of the linear transformations

$$\begin{pmatrix} X_1 \\ P_1 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} X_0 \\ P_0 \end{pmatrix},$$

where $X_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$, $P_i = \begin{pmatrix} p_i \\ q_i \end{pmatrix}$ $(i = 0, 1)$ and the matrices A, B, C and D, each of order two, are such that

$$A^*C - C^*A = 0, \quad B^*D - D^*B = 0, \quad D^*A - B^*C = 1.$$

For this purpose Hamilton's method of characteristic functions is used, and the results are interpreted optically in regard to the formation of images by the narrow bundle, S. When $|C| \neq 0$, it is possible, using the above matrix equation, to express X_0 and X_1, in terms of P_0 and P_1, in the form

$$\begin{pmatrix} X_0 \\ -X_1 \end{pmatrix} = \begin{pmatrix} \tilde{A} + \gamma_0 1, & -F \\ -F, & \tilde{C} + \gamma_1 1 \end{pmatrix} \begin{pmatrix} P_0 \\ P_1 \end{pmatrix},$$

where γ_0 and γ_1 are constants and $\tilde{A}$ and $\tilde{C}$ are symmetric. In the case of an axially symmetric system we obtain the Gaussian first-order optics. Then

$$\tilde{A} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \quad F = \begin{pmatrix} f & 0 \\ 0 & f \end{pmatrix}, \quad \tilde{C} = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}.$$

and by means of these matrices the properties of the Gaussian system are deduced. A generalization is made to the case of a bundle of rays about an oblique meridional ray in a system of revolution. Then

$$\tilde{A} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}, \quad F = \begin{pmatrix} f_1 & 0 \\ 0 & f_2 \end{pmatrix}, \quad \tilde{C} = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix},$$

and we have what is called an orthogonal ray system.

The theory of the third-order Seidel aberrations in a system of axial symmetry is given in Chapter V. The image functions are developed one stage beyond the first order in the neighbourhood of the principal ray. The usual expressions are found for the five image errors, and, in the case of a medium with a continuously variable refractive index, the analogous integral expressions are given. Petzval's theorem on field curvature is discussed and a section deals with chromatic aberrations.

The last chapter deals with the Diffraction Theory of Optical Instruments, that is, the propagation of light waves in a non-homogeneous optical medium. Here we return to the electromagnetic theory. The diffraction of unpolarized light and the diffraction patterns for various types of aberrations are discussed. Problems of resolution also receive attention, including the resolution of objects of periodic structure. Some valuable supplementary notes are given. About forty pages are devoted to the theory of electron optics, and there are some notes by Hertzberger, including information on the optical qualities of glass.

After this survey of the contents of the book little need be said further. It is a pleasure to find that the characteristic function method is emphasized and used extensively. More publicity on behalf of this method might be given if the notes were used as the basis for a printed book more readily available than the mimeographed notes from Brown University.

L. S. GODDARD.

Table of Arc sin x. Pp. xx + 124. $3.50 ; *Tables of Associated Legendre Functions.* Pp. xlvi + 306. $5.00. (Columbia University Press, New York: British agents, Scientific Computing Service, Ltd., 23 Bedford Square, London, W.C.1, 1945.)

These volumes are numbered (17) and (18) in the, by now, well-known series of tables produced by the New York Mathematical Tables Project, directed by Dr. Arnold N. Lowan, and now sponsored by the National Bureau of Standards. They are typical and worthy members of the series.

The purpose of the first volume is to enable 12-decimal values of arc sin x in radian measure to be obtained for any value of x from zero to unity. The function is, of course,

$$\int_0^x \frac{dt}{\sqrt{(1-t^2)}},$$

and since previous volumes have tabulated the transcendental functions (natural logarithm and inverse tangent) arising in the integration of rational algebraic functions, the present table may be regarded as the first instalment of tables of integrals of functions involving the square root of a quadratic. It is believed that work has also been done on the companion functions, arc sinh x and arc cosh x, and it is to be hoped that the publication of similar tables of these functions may be not long delayed.

The main Table gives 12-decimal values of arc sin x at interval 0·0001 from $x=0$ to $x=0·9890$. In this range the greatest attainable accuracy in interpolation is obtained by the use of the second differences, or " modified " second differences, provided. The difficult range, for greater values of x, has been treated as follows. First, the interval has been reduced to 0·00001. Second differences suffice to $x=0·99750$, and modified second differences thence to $x=0·99900$. Fourth differences are given from 0·99900 to 0·99950, but no differences in the remaining range. Here interpolation is to be effected by use of the formula

$$\text{arc sin } (1-v)=\sqrt{(2v)} \cdot f(v),$$

and $f(v)$ is tabulated at interval 0·00001 from $v=0$ to $v=0·00050$. Tables of interpolation coefficients for use with the differences provided are included and worked examples of their use are given in the introduction.

The introduction gives also a full account of the methods of computation and checking adopted and a bibliography. It is clear from the latter that the present Table has more decimals and/or a smaller interval than any of its predecessors, and that it constitutes a fundamental table of great value.

Considering the importance of the Legendre and Associated Legendre functions (e.g., in spherical, ellipsoidal, and toroidal harmonic analysis), it is somewhat surprising that tables of these functions have not been more readily available. The present volume goes far to fill the gap. It tabulates the functions $P_n^m(x)$ and $Q_n^m(x)$ and their first derivatives for integer values of n from 0 to 10 and of m from 0 to 4 (or to n if $n < 4$), and for half-odd-integer values of n from $-1/2$ to $9/2$ and integer values of m from 0 to 4, and for real and pure imaginary values of x. The argument ranges from 1 (in some cases 0) to 10 at interval 0·1. Values of $P_n^m(\cos\theta)$ and $d\{P_n^m(\cos\theta)\}/d\theta$ for θ from 0° to 90° at interval 1° are also given. Normally six significant figures are tabulated. No provision for interpolation is made, but a schedule showing the accuracy attainable by various Lagrangian interpolation formulae precedes each main Table. Some supplementary tables give exact values and normalization factors.

It is clear that this volume will be of very great use to physicists and others whose work requires numerical values of the functions tabulated in it, and the accuracy should be ample for almost all practical applications. W. G. BICKLEY.

Thermodynamics, by PHILIPP FRANK. Pp. iv + 123 + 24. (Brown University, Providence, Rhode Island, 1945.)

These notes, in mimeographed form, were prepared in connection with one of the very valuable summer schools run at Brown University. The scope of the course is, roughly speaking, that covered by the ordinary textbook, restricted to the classical (and perfectly general) thermodynamics ; the statistical approach is not employed. The course, however, is far from being conventional. It is original in being one of the first to adopt Caratheodory's methods, so that, for example, the reader sees clearly that *entropy* is a concept suggested by, and associated with, the first law of thermodynamics, and not, as students commonly suppose, only introduced because of the second law. This fact was, as it happens, known to Kelvin, but it has tended to slip out of sight. In any case, Caratheodory's treatment is of an altogether higher order of rigour than that of any of his predecessors. This is not to say that it uses more difficult mathematics, but its outlook is more critical.

After the bases of the science have been laid, the notes proceed to deal with the applications to pure substances, engines, magnetic and electric phenomena ; the third law is then introduced and its application to problems of chemical equilibrium described. A valuable note on the thermodynamics of gas mixtures is contributed by John A. Goff. In this, the methods of Gibbs are supplemented by recent developments, including the determination of free energies (or equivalent quantities) from spectroscopic data.

Whilst we may envy those who were privileged to attend the summer school, and may find these notes useful and interesting, it would not be true to say that they would make a good book if published as they stand. For that, they would need either to be pruned to make more definitely a set of notes, or else expanded further to give those full explanations which were doubtless added orally when the lectures were delivered.

J. H. A.

Atomic Artillery and the Atomic Bomb. By J. K. ROBERTSON. Pp. viii + 173. (New York: Van Nostrand Co., 1945.) $2.50.

The title of this book appears more sensational than the author really intended, for his artillery is, in fact, alpha particles, protons, electrons and the like. The book is intended for the intelligent layman, and covers roughly the work on atomic physics subsequent to the discovery of the electron by Sir J. J. Thomson. After an explanation of the nature of elements and the significance of atomic weights, the author discusses the concept of the atom, proceeds to the electron and discusses methods of accelerating charged particles. He then mentions the early work on positive rays and the mass spectrograph, gives a short account of radio-activity, and then comes to gamma rays. After this, he deals with the work on cosmic rays which, in the period 1930–35, gave us the new particles of positron and meson (or, as he calls it, the mesotron).

The work with which Rutherford's last period of activity was associated was the transmutation of elements by bombardment with various particles, and the author gives a clear account of the nature of this nuclear chemistry, explaining how both mass number and atomic weight must balance in each reaction. From this point his story becomes quite modern, and we are introduced to fission and chain reactions, and it is explained how these could be used to keep a reaction going, and thus provide the world with a new source of power.

As is now well known, this work led not only to a possible source of power, but incidentally to a means of production of elements not previously known and, in particular, of plutonium. Some short account of this work is given and its application to the atomic bomb is mentioned.

The book, on the whole, seems accurate. There is one slip where *positron* occurs for *proton*, and the reader is led to suppose that Aston is still alive and working, but these seem to be the major slips.

It is satisfactory to read a book on this subject which does not spin words at great length on the sociological and political implications of the work being done, and it is still more satisfactory to know that we can now recommend to our non-scientific friends a book which they can read for themselves and understand in the knowledge that they will not be misled as to the facts.   J. H. A.

Circular and Hyperbolic Functions, Exponential and Sine and Cosine Integrals, Factorial Function and Allied Functions, Hermitian Probability Functions. (*British Association Mathematical Tables, Volume I*). Second edition. Pp. xi+72. (London: Cambridge University Press, 1946.) 10s.

The collection under review is a welcome second edition of the British Association Mathematical Tables, Volume I, originally published in 1931 and now for some years out of print. The functions tabulated (together with even differences, sometimes up to the sixth) are as follows: $\frac{1}{2}\pi x$[1(1)100 ; 15], sin x, cos x [0·0(0·1)50·0 ; 15 and 0·000 (0·001)1·600 ; 11], sinh πx, cosh πx [0·0000(0·0001)0·0100 ; 15 and 0·00(0·01)4·00 ; 15], sinh x, cosh x [0·0(0·1)10·0 ; 15], Ei(x), $-$Ei($-x$) [0·0(0·1)15·0 ; 11], Si(x), Ci(x) [0·0(0·1)20·0(0·2)40·0 ; 11, 10], x ! [0·00(0·01)1·00 ; 12], $\left\{1+\int_0^x \log_{10} t! \, dt\right\}$ [0·00(0·01) 1·00 ; 10], $\frac{d^n}{dx^n}\log_e x$! [$n=1, 2, 3, 4, 0·00(0·01)1·00$ and 10·0(0·1)60·0 ; 10, 12], $Hh_n(x)$ $=\int_x^\infty Hh_{n-1}(x)\,dx$ where $Hh_0(x)=\int_x^\infty e^{-\frac{1}{2}t^2}\,dt$ [$n=0(1)21$, $x=-7·0(0·1)+6·6$; 10], $Hh_0(x)$

$Hh_2(x)/\{Hh_1(x)\}^2$ [$-7·0(0·1)+50·0$; 9 to 3]. In tabular content the new edition thus does not differ from the old, but the opportunity has been taken to provide an auxiliary table extending the values of the Factorial Function, x !, from 12 to 18 places of decimals and to correct eleven errors (of which nine merely affected end-figures by one unit) that have been detected. Possessors of the first edition are reminded that they may obtain a list of these corrections by applying to the Office of the British Association at Burlington House, London W. 1.

A drastic pruning of the new edition of the tables has removed the historical note, and the 30-page introduction which formed such an interesting and informative a feature of the first edition. This loss may not be felt acutely by those who, as it were, already possess a function and merely require its numerical values, but, from the standpoint of general interest and suggestiveness for fruitful application, one experiences a feeling of slight regret at the omission of the material illustrating the analytical properties and uses of the " Polygamma " Functions and the repeated integrals of the Error Function (now called Hermitian Probability Functions).

Mathematical tables, although (when accurate) essentially perdurable, are yet susceptible to loss of pre-eminence by comparison with subsequent more extensive tables. Few users will find the number of significant figures inadequate in the present volume of British Association tables, but, as is obvious from the list of contents given above, for many

purposes subtabulation of several of the functions would be welcome, although inevitably bulky. In this connection it is relevant to remark that since 1931 a new standard for the close tabulation of a large number of functions has been established by the publications of the New York Mathematical Tables Project.

Criticism of a book because it is not a whole library is, however, unfair. The volume under review contains a diverse, interesting and most useful collection of tables which yield their full value to those prepared to interpolate accurately. The price is very moderate, despite the fact that the present binding is not sufficiently stout for its hard future. It is a pleasure to consult tables so elegantly set and printed. M. S. JONES.

The Music Review, Vol. VII, No. 1. (Heffer & Sons, Cambridge.) 5s.

This number contains an article by B. van der Pol on " Music and Elementary Theory of Numbers ", which deals with time and rhythm, absolute and relative pitch, Euler's theory of dissonance, musical scales, relaxation oscillations and variations of musical pitch. The author's investigations on the variation in pitch between orchestras which broadcast was described in Vol. VII of the *Progress Reports* (pp. 33–34). The remaining topics are considered either in Helmholtz's *Sensations of Tone* or in modern textbooks of sound, but the musical physicist will find in this article a useful summary of present knowledge in this field, and also some useful observations on the sense of absolute pitch, studied by the author, himself a musician with this gift and with the scientist's sense of impartiality. E. G. R.

Statistical Thermodynamics, by ERWIN SCHRÖDINGER. Pp. 88. (Cambridge: The University Press, 1946). 6s.

This very small book contains a great deal of material, which was originally presented as a course of lectures at the Dublin Institute for Advanced Studies, in the spring of 1944. It is hardly a reference work to which a man would turn when presented with a specific problem which he wished to solve, but rather a background exposition which he should read and ponder when he is not busy applying statistical methods to particular problems.

Schrödinger adopts the attitude that there is only one problem in statistical thermodynamics, that of determining the distribution of an amount E of energy over N similar systems, which may be only mental copies of one real system. From this starting point he considers how the distributions may be enumerated, and in doing so he is led to discuss the logical bases of the procedure, and also to sketch the mathematical apparatus normally used. He is more interested in the difficulties of the theory than in those parts where all is straightforward, and consequently has written a most stimulating book, which is well worth its price. Incidentally, though produced under war-time conditions, the book is pleasant to handle and to look at—a tribute to the skill of the printers. Among matters which are discussed are the setting equal to zero of the constant entropy which remains in a crystalline system at absolute zero, and the relationship between the classical, the Bose-Einstein and the Fermi-Dirac statistics. All these are subjects on which a great deal has been written before, but the author's peculiar quality of illuminating those matters on which he writes makes his treatment very well worthy of study. J. H. A.

CORRIGENDUM

" Crystal structure of double oxides of the perovskite type ", by HELEN D. MEGAW (*Proc. Phys. Soc.*, 58, p. 133).

The address given under the author's name should read " Philips' Lamps, Limited, Mitcham, Surrey ; now at Birkbeck College, London ".

LIGHT EMISSION DURING COOLING OF A PLANCKIAN RADIATOR

BY E. F. CALDIN,

Department of Physical and Inorganic Chemistry, Leeds University

ABSTRACT. General expressions are derived for the variation of luminous intensity with time, and the corresponding integral of the intensity with respect to time, for a Planckian radiator of uniform temperature, cooling by radiation from temperatures above 2000° K. The time-intensity integral for light emitted in a given direction is found to depend only on the initial temperature, heat capacity and shape of the body ; the form of the intensity-time curve depends also on the surface area and emissivity. Values of the time-intensity integral for a spherical total radiator of unit heat capacity, cooling from various temperatures, are tabulated. The relevance of these results to the study of certain light sources of short duration is indicated.

§1. INTRODUCTION

THE study of transient light sources of short duration presents special problems because of the rapid continuous variation of the luminous intensity of the source. Examples of such sources are: (*a*) metal wires burning in oxygen (Van Liempt and de Vriend, 1934, 1935); (*b*) photographic flash powders, such as the Agfa powder, consisting of a mixture of magnesium and thorium nitrate (Beck and Eggert, 1926); and (*c*) the photoflash bulb used in photography, such as the Sashalite bulb, in which aluminium foil burns in oxygen (Reeb, 1935; Van Liempt and de Vriend, 1937). The luminous intensity of such a source increases rapidly to a maximum and then exhibits a much slower decrease which is roughly exponential with time (Beck and Eggert, 1926; Reeb, 1935; Van Liempt and de Vriend, 1937; Kelley, 1938; Projector and Barbrow, 1945). These sources exhibit continuous spectra (cf. following paper) and may well consist essentially of incandescent solid particles—the oxide produced by combustion of the metal—raised to incandescence by the heat of combustion, and cooling by radiation. It is thus of interest to compare the declining part of the intensity-time curve with the corresponding curves for Planckian radiators, initially at a high temperature, and cooling by radiation. Further, the actinic effect of such a source, i.e. the density of a photographic plate exposed to it under given conditions, will depend (neglecting reciprocity-law failure) not on the maximum intensity but on the integral with respect to time of the intensity, i.e. the area under the intensity-time curve; so that the value of this integral for a Planckian radiator is also of interest. Experimental values for the time-intensity integral for the above sources have been given by Beck and Eggert (1926), Eder (1930), Reeb (1935), and Van Liempt and de Vriend (1934, 1935).

In this paper, general expressions are derived for the variation of luminous intensity with time, and for the integral of the intensity with respect to time, for Planckian radiators cooling from temperatures above 2000° K. The derivation makes use of the relation already reported between the brightness and temperature of a total radiator (Caldin, 1945). To obtain manageable expressions, the temperature of the cooling body has been assumed uniform and temperature gradients ignored. A table of values is given for the time-intensity integral for a spherical total radiator of unit heat capacity, cooling from various temperatures; this allows easy calculation of the value of the integral for other simple cases. The conditions necessary for the valid application of the theory to data on actual sources are being examined by Dr. F. R. Booth. None of the published experimental work is complete enough to allow a satisfactory test of the theory; some new experimental work will be presented, it is hoped, in a subsequent paper.

The question of units of intensity arises. Calculations using the Planck equation and the standard C.I.E. relative luminosity data give values of $\int_0^\infty V_\lambda J_\lambda d\lambda$. Here V_λ is the value of the C.I.E. relative luminosity factor at the wave-length λ; $J_\lambda d\lambda$ is the total energy flux from unit area of the total radiator between the wave-lengths $\lambda \pm d\lambda/2$, as given by Planck's equation. If $J_\lambda d\lambda$ is expressed in watts per cm^2, the luminous flux per unit area is then given in lightwatts per cm^2, according to the definition of Moon (1936). It would be very convenient if corresponding units were in use for intensity, brightness, etc., as well as for luminous flux. The brightness of a total radiator would be given in these units by $\frac{1}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$; and the intensity in a given direction would be $\frac{A'}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$, where A' is the projected area of the total radiator on a plane perpendicular to the given direction. However, such physically-defined units are not in use. In a previous paper (Caldin, 1945), the values of $B = \frac{1}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$ for various temperatures were calculated, and given as values of the brightness of a total radiator expressed in lightwatts per cm^2 of projected area (more accurately, lightwatts cm^{-2} steradian^{-1}). But this is not a recognized practice among photometrists, and in order not to raise here the whole question of physical definition of photometric units we shall use the international candle as the unit of intensity, and the lumen as the unit of luminous flux (cf. *International Lighting Vocabulary*, C.I.E. 1938). The ratio of the lumen to the lightwatt, M, will therefore appear in the equations. Thus, the total luminous flux from unit area of a total radiator is $M\int_0^\infty V_\lambda J_\lambda d\lambda$ lumens; the brightness B_c is $\frac{M}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$ candles per cm^2 of projected area ; and the intensity in a given direction is $\frac{A'M}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$ candles, where A' is the projected area of the source on a plane perpendicular to that direction. The ratio M is somewhat uncertain, but we shall not need to insert a numerical value for it for most of the purposes of this paper. Values of M for various values of the radiation constants, and a general method of computation, have been given in a previous paper (Caldin, 1946).

The term *Planckian radiator* is used in this paper to denote a radiator whose emissivity is constant with respect to wave-length and temperature, though not necessarily equal to unity. The energy flux from such a body at any wave-length and temperature bears a constant ratio to that from a total radiator at the same wave-length and temperature. The term thus covers both "black" and "grey" bodies.

§2. DERIVATION OF EXPRESSIONS FOR LIGHT EMISSION DURING COOLING OF A PLANCKIAN RADIATOR

We consider a Planckian radiator, cooling by radiation alone. Its temperature will be assumed uniform. The emissivity of the body will be assumed to be independent of temperature and wave-length, and the specific heat independent of temperature, at the relevant temperatures. The surface is assumed to be a perfect diffuser.

Let

m = mass of body (gm.)

A = surface area of body (cm².).

s = specific heat of body (cals. gm.⁻¹ deg. ᴋ⁻¹).

E = emissivity.

t = time from start of cooling, in seconds.

H = total heat content of body, at time t sec. (cals.).

T = temperature of body at time t sec. (°ᴋ.).

T_0 = initial temperature of body at $t=0$.

T_s = temperature of surroundings, assumed constant, and $< T_0$.

$B = \dfrac{1}{\pi}\displaystyle\int_0^\infty V_\lambda J_\lambda d\lambda$, where $J_\lambda d\lambda$ is in watts cm.⁻²

B_e = brightness of body = luminous intensity per unit projected surface area, in a given direction, in candles cm.⁻² $= MB = \dfrac{M}{\pi}\displaystyle\int_0^\infty V_\lambda J_\lambda d\lambda.$

M = ratio of lumen to lightwatt.

I = luminous intensity, in a given direction (candles).

$W = \displaystyle\int_0^\infty B_e dt$ = integral with respect to time of the intensity per unit projected area in a given direction; the integral being taken over the whole duration of the cooling (candle-sec.).

K = Stefan's constant.

$k = KAE/ms.$

By Stefan's law we have, then, for the net rate of energy loss from the body by radiation,

$$-dH/dt = KAE(T^4 - T_s^4). \qquad \ldots\ldots(1)$$

We have, also, since the temperature of the body is assumed to remain uniform,

$$dH = ms\,dt. \qquad \ldots\ldots(2)$$

Substituting for dH in equation (1),

$$-dT/dt = \frac{KAE}{ms}(T^4 - T_s^4),$$

or, writing $$k = KAE/ms,$$

$$-dT/dt = k(T^4 - T_s^4).\qquad\qquad\ldots\ldots(3)$$

This may be integrated by partial fractions, giving

$$t = \frac{1}{4kT_s^3}\left[12\,t\text{an}^{-1}\,T/T_s + \ln\,(T + T_s)/(T - T_s)\right]_{T_0}.\qquad\ldots\ldots(4)$$

On expanding the two terms in the bracket as series, and inserting the limits T_0 and T, we obtain the following relation:

$$t = [(T^{-3} - T_0^{-3})/3k] + [T_s^4(T^{-7} - T_0^{-7})/7k] + \ldots\qquad\ldots\ldots(5)$$

All the terms of this equation after the first may generally be neglected when T_s is of the order of room temperature; thus, for example, when $T_0 = 2640°\,\text{K.}$, $T = 2000°\,\text{K.}$, and $T_s = 300°\,\text{K.}$, the second term is only $1/3000$ of the first, and later terms will be still smaller. Thus, as a very good approximation, we may write

$$T = (3kt + T_0^{-3})^{-\frac{1}{3}}.\qquad\qquad\ldots\ldots(6)$$

From this relation between the temperature and time, a relation between the brightness and time is found as follows. In a previous paper (Caldin, 1945) values of B $\left(\text{i.e.}\ \dfrac{1}{\pi}\displaystyle\int_0^{\infty} V_\lambda J_\lambda d\lambda\right)$ were given for a total radiator at a series of temperatures from $2000°\,\text{K.}$ to $3120°\,\text{K.}$; these were calculated by combining Planck's energy-wave-length-temperature relation with the C.I.E. standard luminosity function. It was found that B could be expressed as a function of temperature (T) by an equation of the form

$$\log_{10} B = \alpha - \beta/T,\qquad\qquad\ldots\ldots(7)$$

where α and β are constants which depend on the units used for B and T and on the values of the constants C_1 and C_2. Using Birge's 1941 values for C_1 and C_2 ($C_1 = 3{\cdot}7403 \times 10^{-5}$ erg. sec.$^{-1}$ cm.$^{+2}$, $C_2 = 14384{\cdot}_8$ micron degrees), and expressing T in degrees K., the values of the constants are: $\alpha = 4{\cdot}279$, $\beta = 10902$. (With the values $C_1 = 3{\cdot}703 \times 10^{-5}$, $C_2 = 14330$ (*International Critical Tables*), we find $\alpha = 4{\cdot}275$, $\beta = 10860$.)

For a source of emissivity E (assumed constant with respect to temperature and wave-length), the equation corresponding to (7) can be written

$$B = aE \exp(-c/T),\qquad\qquad\ldots\ldots(8)$$

where $\log_{10} a = \alpha$, and $c = \beta/0{\cdot}43429$. With Birge's values of C_1 and C_2, $c = 25103$ when T is in $°\,\text{K.}$, and $a = 1{\cdot}90_1 \times 10^4$. These values will be referred to as c_1 and a_1.

Hence for the brightness, expressed in candles per cm.^{2}, we obtain

$$B_c = MB = aEM \exp(-c/T).\qquad\qquad\ldots\ldots(9)$$

A relation between the brightness of a cooling total radiator initially at temperature T_0, and the time t which has elapsed since cooling began, is obtained by combining equations (6) and (9). The result is

$$B_c = aEM \exp\left[-c(3kt + T_0^{-3})^{\frac{1}{3}}\right].\qquad\qquad\ldots\ldots(10)$$

The definite integral of this expression with respect to time from $t = 0$ to $t = \infty$, corresponding to the temperature limits T_0 to T_s, is:

$$W = \int_0^\infty B_c \, dt = \frac{aEM}{kc^3} [(u^2 + 2u + 2) \exp(-u)], \qquad \ldots \ldots (11)$$

where $u = c/T_0$. This integration presupposes that the relation (9) holds with sufficient accuracy down to T_s, which may be well below $2000°$ K., the lower limit for the calculated values of B on which equation (9) is based. We shall later estimate the accuracy and show that it is in fact sufficient when T_0 is of the order of $3000°$ K.

On replacing k by KAE/ms, and writing $f(u)$ for $[(u^2 + 2u + 2) \exp(-u)]$, and G for $a/c^3 K$., equation (11) becomes

$$W = MG(ms/A)f(u). \qquad \ldots \ldots (12)$$

If we insert the values given above for a and c, i.e. a_1 and c_1, corresponding to Birge's 1941 values of C_1 and C_2, and his value of K, namely $1 \cdot 354 \times 10^{-12}$ cal. cm.$^{-2}$ deg.$^{-4}$ sec.$^{-1}$, we find $G = 887$ (dimensions: $°$K.). This value will be referred to as G_1. (With $C_1 = 3 \cdot 703 \times 10^{-5}$ and $C_2 = 14330$, we find $G = 889$.)

Equations (9) and (12) both refer to the luminous flux in a given direction due to unit projected area of the radiating source. To obtain the luminous intensity (in candles) in a given direction due to the whole body, at a given time, and the integral with respect to time of the intensity (in candle-sec.), we must multiply B_c and W respectively by A', the projected area of the body on a plane perpendicular to the given direction.

Thus finally, for the *intensity at time t* we obtain

$$I = A'B_c = aEA'M \exp[-c(3kt + T_0^{-3})^{\frac{1}{3}}], \qquad \ldots \ldots (13)$$

where $I =$ luminous intensity in candles and the other symbols have already been defined.

From (13) the *initial (maximum) intensity* is seen to be

$$I_{\max} = aEA'M \exp(-c/T_0) = aEA'M \exp(-u). \qquad \ldots \ldots (14)$$

and the *intensity at time t, relative to the maximum*, is

$$I/I_{\max} = \exp[c/T_0 - c(3kt + T_0^{-3})^{\frac{1}{3}}]. \qquad \ldots \ldots (15)$$

For the *time-integral of the intensity* we obtain

$$\int_0^\infty I \, dt = A' \int_0^\infty B_c \, dt = A'W = MG \cdot ms(A'/A)f(u). \qquad \ldots \ldots (16)$$

Thus the integral of the intensity in a given direction with respect to time $(A'W)$—which is the main factor controlling the total actinic effect of such a source as we are considering—depends, according to equation (16), on the initial temperature (since $c/u = T_0$), the heat capacity (ms), and the ratio A'/A, which depends on the shape of the body (for a sphere it is 1/4). It does not depend on the emissivity. The initial temperature has the greatest proportional influence; whereas the value of $A'W$ is directly proportional to the heat capacity and to A'/A, a small change of initial temperature in the region of $2500°$ K. (for example) produces a change in $A'W$ proportionally about eight times greater.

The initial intensity in a given direction, I_{max}, depends, according to equation (14), on the emissivity (E), projected area (A'), and initial temperature. The shape of the intensity-time curve depends, according to equation (15), on the emissivity, total surface area, heat capacity (since $k = KAE/ms$), and initial temperature; the time required for the intensity to fall to a given fraction of its initial value, for a given value of T_0, is proportional to $1/k$, and, therefore, increases with (a) decrease of emissivity, E; (b) decrease of specific surface area, A/m; and (c) increase of specific heat, s.

If we are considering the light emitted by a body whose mass, dimensions and specific heat are known, but whose initial temperature and emissivity are unknown—as generally when we are dealing with the experimental data on these transient sources—we can proceed as follows. The observed value of the intensity-time integral, with the mass, specific heat and shape of the body, gives by equation (16) the value of $f(u)$, and hence of the initial temperature T_0. The observed initial intensity then gives the emissivity E, by equation (18). Then the whole intensity-time curve, or relative intensity-time curve, can be calculated from equation (13) or (15), and compared with the observed curve. For a source of given heat capacity and dimensions, a choice of values for k and T_0 is enough to fix the whole theoretical intensity-time curve (and its integral), since the value of E corresponding to the chosen values of k is at once calculable from the relation $k = KAE/ms$, and all the required constants are then known.

§ 3. NUMERICAL VALUES FOR TIME-INTENSITY INTEGRAL

In the table below are given the calculated values of $f(u) = f(c/T_0)$ at a series of temperatures from 2000° K. to 6000° K., with the value c_1 for c. Numerical

Table 1

T (°K.)	$f(c_1/T_0)$	$(671 G_1 f(c_1/T_0)/4) \times 10^3$
2000	0·000654	0·097
2080	0·000984	0·146
2160	0·00144	0·214
2240	0·00204	0·303
2320	0·00281	0·418
2400	0·00379	0·564
2480	0·00501	0·745
2560	0·00650	0·966
2640	0·00826	1·23
2720	0·0104	1·54
2800	0·0128	1·91
2880	0·0156	2·33
2960	0·0188	2·80
3040	0·0225	3·34
3120	0·0265	3·95
4000	0·0977	14·5
5000	0·233	34·6
6000	0·394	58·6

values of the time-intensity integral may be obtained by inserting these values in equation (16). In the table are given also values of $671 G_1 f(u)/4$, which, as may

be seen from equation (16), is the value obtained for the time-intensity integral, in candle-seconds, for light emitted in any direction from a spherical total radiator of unit heat capacity, cooling from a given temperature T_0, if we adopt the value G_1 for G and the corresponding value 671 for M (Caldin, 1946).

§4. ACCURACY OF FORMULAE (13) TO (16)

Accuracy when T_0 is less than 3120° K.

It remains to consider the accuracy of the values of I and of $\int_0^\infty I\,dt$ calculated from equations (13) to (16). The percentage error in I is equal to the percentage error in the value of B calculated from equation (8), with the values of a and c assumed above; this, as already stated, is not more than 1%. The error in W is made up of an error equal to this and an error due to the assumption that the relation (8), which has been tested by calculation only down to 2000° K., holds also between 2000° K. and the final temperature T_s, with the same values of a and c,

Actually the relation (8), with the constants a_1 and c_1 given above, begins to deviate appreciably from the truth below 2000° K. This may be shown by

Table 2

T (°K.)	$\log_{10} B$ (Ives) (a)	$\log_{10} B$ from eqn. (17) (b)	Difference (a)−(b)	$\log_{10} B$ from eqn. (22) (c)	Difference (a)−(c)
2000	$\bar{2}{\cdot}854$	$\bar{2}{\cdot}845$	+0·009	$\bar{2}{\cdot}853$	+0·001
1900	$\bar{2}{\cdot}572$	$\bar{2}{\cdot}559$	+0·013	$\bar{2}{\cdot}573$	−0·001
1800	$\bar{2}{\cdot}260$	$\bar{2}{\cdot}242$	+0·018	$\bar{2}{\cdot}262$	−0·002
1700	$\bar{3}{\cdot}914$	$\bar{3}{\cdot}887$	+0·027	$\bar{3}{\cdot}915$	−0·001
1600	$\bar{3}{\cdot}525$	$\bar{3}{\cdot}487$	+0·038	$\bar{3}{\cdot}524$	+0·001
1400	$\bar{4}{\cdot}591$	$\bar{4}{\cdot}518$	+0·073	$\bar{4}{\cdot}575$	+0·016
1200	$\bar{5}{\cdot}356$	$\bar{5}{\cdot}225$	+0·131	$\bar{5}{\cdot}310$	+0·046

comparing values of B calculated from it with the values calculated by Ives (1926). Ives used the C.I.E. relative luminosity data and the value 14330 micron degrees for C_2; however, instead of using the actual C.I.E. values, he used a formula which reproduced them with an error of the order of 1%; his values of B are subject to the same error. He gives values of $B_c = \dfrac{M}{\pi} \int_0^\infty V_\lambda J_\lambda \, d\lambda$ with the value $M = 1/0{\cdot}00161$. For comparison with the results of equation (8), the corresponding values of $B = B_c/M$ have been computed. These values are compared in the following table with those calculated from the equation

$$\log_{10} B = 4{\cdot}275 - 10860/T, \qquad \ldots\ldots(17)$$

in which the constant 10860 corresponds with $C_2 = 14330$ micron degrees. The fourth column gives the difference in log units.

It is clear that the divergence between the directly calculated values and those derived from equation (17) increases with decrease of temperature; it becomes

as much as 9% at 1600°K. Since, however, very little light is emitted at such low temperatures, the error in the time-intensity integral will be much less than this. This error is best estimated as follows. Ives' values from 2000° K. down to 1600° K. are represented within 0·2% by the equation

$$\log_{10} B = 4\cdot168 - 10630/T, \qquad \ldots\ldots(18)$$

as may be seen from the sixth column of the table above. We can thus use equation (18) below 2000°K.; down to the lowest temperature which need be considered, with the same accuracy as that afforded by equation (17) between 2000°K. and 3120°K. Correcting equation (18) to correspond with Birge's 1941 values of C_1 and C_2, we get for temperatures below 2000° K.

$$\log_{10} B = 4\cdot172 - 10670/T, \qquad \ldots\ldots(19)$$

and the constants in equation (8) become $a = 1\cdot49 \times 10^4$, $c = 24568$. This gives G in equation (16) as 740. Let these values of the constants be designated a_2, c_2, G_2, and the values found for the temperature range 2000° to 3120°K. be a_1, c_1, G_1. Then the value of the time-intensity integral for light emitted below 2000°K. is, by equation (16), $[G_2 ms(A'/A)f(c_2/2000)]$. The corresponding value obtained by assuming the values of the constants a_1, c_1 and G_1 to hold over the whole temperature range T_0 to T_s is $[G_1 ms(A'/A)f(c_1/2000)]$. Thus the error in using a_1, c_1 and G_1 below 2000° K. would be

$$[ms . A'/A][G_1 f(c_1/2000) - G_2 f(c_2/2000)] \qquad \ldots\ldots(20)$$

and the percentage error in the whole value of the time-intensity integral for a body cooling from a temperature T_0 would be

$$100\,[f(c_1/2000) - \frac{G_2}{G_1} f(c_2/2000)]/f(c_1/T_0). \qquad \ldots\ldots(21)$$

Inserting numerical values, this gives

$$\text{percentage error} = 0\cdot003/f(c_1 T_0). \qquad \ldots\ldots(22)$$

Thus the percentage error depends only on T_0, not on k; and, as would be expected, it decreases with increase of T_0. For $T_0 = ca. 3000°$ K., $f(c_1/T_0)$ has the value $ca. 0\cdot02$, so that the time-intensity integral for light emitted below 2000° K. is about 3% of the total (this fraction depends only on T_0), and the error due to the use of the values a_1 and c_1 throughout is 0·15%. The total error in W would then be less than 1%.

Accuracy with higher values of T_0.

So far, only values of T_0 lower than 3120°K. have been considered. To estimate the additional errors in using formula (13) to (16) with higher values of T_0 and the same values a_1, c_1 and G_1 for the constants, values of B calculated from equation (17) may be compared with those derived from Ives' calculated values of B_c for higher temperatures, as in table 3.

Thus the error in calculating B from equation (17) is under 1% with $T_0 = 3000°$K.; 3% with $T_0 = 4000°$; 7% with $T_0 = 5000°$, and 9% with $T_0 = 6000°$K. The percentage errors in calculating I from equation (13), or I_{max} from equation (14), will correspond with these errors. The percentage error in the time-intensity integral will be less; we can estimate it by a method

somewhat similar to that of the previous section. Values of B derived from Ives' values of B_c from 3000° K. to 6000° K. can be represented within about $\pm 1\%$ by

$$\log_{10} B = 4 \cdot 339 - 11040/T \qquad \ldots\ldots(23)$$

in place of equation (17) or (18). Converting the constants to correspond with Birge's 1941 values of C_1 and C_2, we find that the corresponding values of the

Table 3

T (°K.)	$\log_{10} B$ (Ives) (a)	$\log_{10} B$ from eqn. (17) (b)	Difference (a)−(b)
3000	0·659	0·655	+0·004
4000	1·574	1·560	+0·014
5000	2·131	2·103	+0·028
6000	2·504	2·465	+0·039

constants a, c and G are $a_3 = 2 \cdot 20_3 \times 10^4$, $c_3 = 25518$, $G_3 = 980$. It may then be easily shown that the additional percentage error due to using a_1, c_1, and G_1 in calculating the time-intensity integral is

$$\frac{100[G_1 f(c_1/T_0) - G_3 f(c_3/T_0) - G_1 f(c_1/3000) + G_3 f(c_3/3000)]}{[G_1 f(c_1/3000) - G_3 f(c_3/3000) + G_3 f(c_3/T_0)]} . \qquad \ldots\ldots(24)$$

Inserting numerical values, we find, for $T_0 = 4000°$ K., $-2 \cdot 6\%$; for 5000° K., $-4 \cdot 3\%$; for 6000° K., $-5 \cdot 4\%$.

ACKNOWLEDGMENT

REFERENCES

BECK and EGGERT, 1926. *Z. wiss. Photogr.*, **24**, 376.
BIRGE, R. T., 1941. *Rev. Mod. Phys.*, **13**, 233.
CALDIN, E. F., 1945. *Proc. Phys. Soc.*, **57**, 440 ; 1946. *Proc. Phys. Soc.*, **58**, 207.
EDER, J. M., 1930. *Z. wiss. Photogr.*, **27**, 337.
HYDE, FORSYTHE and CADY, 1919. *Phys. Rev.*, **13**, 45.
IVES, H. E., 1926. *J. Opt. Soc. Amer.*, **12**, 75.
VAN LIEMPT and DE VRIEND, 1934. *Rec. Trav. Chim. Pays-Bas*, **53**, 839 ; 1935. *Ibid.*, **54**, 239 ; 1937 : *Physica*, **4**, 353.
KELLEY, 1938. *J. Opt. Soc. Amer.*, **28**, 27.
MOON, P., 1936. *Scientific Basis of Illuminating Engineering* (McGraw-Hill), pp. 53, 57, 552.
PROJECTOR, T. H. and BARBROW, L. E., 1945. *Rev. Sci. Instrum.*, **16**, 51.
REEB, 1935. *Z. wiss. Photogr.*, **34**, 77.

THE DETERMINATION OF THE INITIAL TEMPERATURE OF A COOLING TOTAL RADIATOR FROM MEASUREMENTS OF THE SPECTRAL DISTRIBUTION OF THE ENERGY EMITTED DURING COOLING

By E. F. CALDIN,

Department of Physical and Inorganic Chemistry, Leeds University

ABSTRACT. An expression is derived theoretically relating the initial temperature (T_0) of a total radiator, cooling by radiation, to the spectral distribution of the total energy emitted during cooling (i.e. the variation with wave-length of the integral with respect to time of the energy flux at a given wave-length). This expression allows the determination of the initial temperature of such a cooling source (to which photographic flash bulbs, etc., provide approximations), from the spectral distribution of the energy received by a photographic plate or other recording device exposed to the source. This spectral distribution is compared with that of the energy flux from a total radiator at various temperatures ; it is found that for each value of T_0 (between 2500° K. and 4500° K.) a temperature (T_1) can be found such that there is close agreement between the two spectral distributions. T_1 corresponds to the " temperature " cited in the literature for such sources. A linear relation is found to hold between T_1 and T_0.

§1. INTRODUCTION

I N a previous paper (Caldin, 1946 b), expressions were derived for the time-intensity relation, and the corresponding time-intensity integral, for light emission from a total radiator of uniform temperature, cooling by radiation. The calculations are of interest because the results can be compared with those found experimentally for certain light sources which give continuous spectra and may owe their emission to cooling solid incandescent particles. Examples include certain sources of short duration, such as those produced by combustion of metal wires or foil in oxygen (as in photographic flash bulbs), or by combustion of photographic flash powders; and also some continuous sources, e.g. the "Nitralamp", a projection lamp in which magnesium ribbon is burned. Besides parameters such as the mass, shape and emissivity of the source, the expressions derived contain T_0, the initial temperature. This is not readily accessible, though in principle it might be obtained by continuously recording, by means of cathode-ray oscillographs, the relative intensities of the light in, say, the red and blue regions of the spectrum. Values given for the temperatures of these sources have usually been obtained by burning the flash powder, etc., in front of the slit of a spectrograph, exposing the plate to the flash during the whole of its duration, and comparing the distribution of density along the spectrogram with

the density-distributions produced by sources emitting approximately Planckian radiation at various constant temperatures. It is evident that, since the temperature is falling continuously, the spectral distribution of the total energy which falls on the plate (i.e. the spectral distribution of the integral of the energy flux with respect to time) will not be the same as the spectral distribution corresponding to the initial temperature of the source, and the derived "temperature" will be lower than T_0. The following values for the "temperatures" of such sources have been recorded:—

Source.	"Temperature" ($^{\circ}$ K.)	Reference.
Aluminium foil in oxygen (photoflash lamp)	3500	Forsythe and Easley (1934)
Magnesium ribbon in air	3700 3750	Dziobek (1928) ; O'Brien and Russell (1933)
Magnesium ribbon in oxygen	*ca.* 4000	Eder (1930)
Thorium nitrate-magnesium mixture (equal proportions) = Agfa flash powder.	3000–3100 3350	Eder (1930) ; Arens and Eggert (1930)

It has not, however, been shown that the radiation from such sources is strictly Planckian; indeed, the likelihood of variation of emissivity, and of emission of band spectra from AlO and MgO, make this very doubtful.

In this paper, a relation is sought between the initial temperature T_0 of a total radiator, which cools by radiation alone, and in which temperature equilibrium is maintained, and the "temperature", T_1, derived from the spectral distribution of the time integral of the energy flux. The latter distribution is calculated in terms of T_0, using the time-temperature relation given in the preceding paper. It is found that the distribution is determined by T_0 alone and does not depend on any other of those properties of the source which control its rate of cooling. It does not correspond exactly with the spectral distribution of the energy flux from a Planckian radiator at any single temperature T_1; but a value of T_1 can be found such that the spectral distribution of the energy flux of a total radiator at T_1° K. agrees quite closely with the spectral distribution of the time integral of the energy flux from a cooling radiator initially at some higher temperature T_0° K. over the whole visible range, at any rate when T_0 is below 4000° K. Values of T_1, for various values of T_0, are obtained, and a relation between T_1 and T_0 derived. It is thus possible, from any given value of T_1, to find T_0, the initial temperature; no knowledge of the emissivity, area, specific heat, etc., of the source is required, provided only that the emissivity is constant with respect to temperature and wave-length.

Work is in progress in several laboratories to determine the spectral distribution both of the time-integral of the energy-flux and of the energy-flux at the maximum temperature, for the light emitted by aluminium foil burning in oxygen. The calculations of this paper should be of assistance in interpreting the experimental results.

§2. CALCULATION OF SPECTRAL DISTRIBUTION OF TOTAL ENERGY
(INTEGRAL OF ENERGY FLUX WITH RESPECT TO TIME)
EMITTED DURING COOLING OF A TOTAL RADIATOR

The spectral distribution of the energy-flux from a total radiator at temperature T is given by Planck's equation:

$$J_\lambda = \frac{C_1 \lambda^{-5}}{e^{C_2/\lambda T} - 1}, \qquad \ldots\ldots(1)$$

where C_1 and C_2 are constants, and $J_\lambda d\lambda$ is the total energy-flux from unit area of a total radiator (watt cm.$^{-2}$) between the wave-lengths $\lambda \pm \dfrac{d\lambda}{2}$ (microns). The simplified form of this equation, due to Wien, holds for wave-lengths up to 0.7μ within (for example) about 0·3 % at 3000° K., and within about 2 % at 5000° K:

$$J_\lambda = C_1 \lambda^{-5} \exp\left(-C_2/\lambda T\right). \qquad \ldots\ldots(2)$$

The relation between temperature and time, for a total radiator which cools by radiation alone, and in which temperature equilibrium is maintained, is given to a close approximation by the following equation (Caldin, 1946)

$$T = (3kt + T_0^{-3})^{-1/3} \qquad \ldots\ldots(3)$$

where

T = temperature at time t (° K.); T_0 = temperature at zero time;

$k = KAE/ms$, where

K = Stefan's constant;

A = surface area of the radiating body (cm.^{2});

E = emissivity of body, assumed constant with respect to wave-length and temperature;

m = mass of body (gm.); and

s = specific heat of body, assumed constant with respect to temperature (cal. gm.$^{-1}$ deg. K.$^{-1}$).

Substituting in equation (2) the value of T given by equation (3), we find for the variation with time of the energy flux per cm.^{2} per micron at a given wave-length

$$J_\lambda = C_1 \lambda^{-5} \exp\left[-\frac{C_2}{\lambda}(3kt + T_0^{-3})^{1/3}\right]. \qquad \ldots\ldots(4)$$

The definite integral of this expression with respect to time, from $t = 0$ to $t = \infty$, is

$$\int_0^\infty J_\lambda dt = \frac{C_1 \lambda^{-4}}{C_2 k T_0^2}\left[1 + \frac{2\lambda T_0}{C_2} + 2\left(\frac{\lambda T_0}{C_2}\right)^2\right] \exp\left(-C_2/\lambda T_0\right). \qquad \ldots\ldots(5)$$

This is the required expression relating the wave-length to the integral with respect to time of the corresponding energy-flux per micron per cm.^{2}

It may be compared with equation (2) and we see that whereas, at a given temperature T_0, the variation J_λ with wave-length is given by the expression

$$J_{\lambda, T_0} \propto [\lambda^{-5} \exp\left(-C_2/\lambda T_0\right)], \qquad \ldots\ldots(6)$$

the variation with wave-length of the time integral of J_λ, for cooling from T_0, is given by an expression containing an extra term:

$$\int_0^\infty J_\lambda dt \propto \left[\lambda^{-5} \exp\left(-C_2/\lambda T_0\right)\right]\left[\lambda\left\{1 + \frac{2\lambda T_0}{C_2} + 2\left(\frac{\lambda T_0}{C_2}\right)^2\right\}\right].$$
$$\dots\dots(7)$$

It is convenient to express the spectral distribution with respect to unity at some wave-length. Thus, taking the values relative to that at $\lambda = \lambda_1 = 0.59\,\mu$, and combining (6) and (7), we obtain:

$$\int_0^\infty J_\lambda dt \bigg/ \int_0^\infty J_{0.59}\, dt = [J_{\lambda,\,T_e}/J_{0.59,\,T_e}]\,[\lambda(1 + 2v + 2v^2)/\lambda_1(1 + 2v_1 + 2v_1^2)],$$
$$\dots\dots(8)$$

where $$v = \lambda T_0/C_2, \qquad v_1 = \lambda_1 T_0/C_2.$$

The values of the second bracket in the right-hand side of equation (8) are a measure of the difference between the spectral distribution of the energy radiated, to a bolometer or a photographic plate, from a total radiator maintained at temperature T_0, and that of the integral with respect to time of the energy flux from the same radiator while cooling from T_0. It will be seen that the latter energy-distribution depends only on the initial temperature, T_0, and not on the parameter k, which is also concerned in determining the rate of cooling. Thus for any value of T_0 we can plot curves of $\left(\int_0^\infty J_\lambda\, dt \big/ \int_0^\infty J_{0.59}\, dt\right)_{T_e}$ against λ, which will be valid for all total radiators, provided that the assumptions implied in the use of equation (3) are fulfilled. These curves may be compared with the curve of $(J_\lambda/J_{0.59})_T$ against λ, i.e. with the spectral distribution of the energy-flux at a given temperature, T, for various values of T. If the two spectral distributions agree fairly closely for some value of T, say T_1, then T_1 is the "temperature" of the transient source obtained according to the methods mentioned in the introduction.

§3. RESULTS AND DISCUSSION

Calculations of $\left(\int_0^\infty J_\lambda dt \big/ \int_0^\infty J_{0.59} dt\right)_{T_e}$ have been carried out for the following values of T_0:—2500°, 3000°, 3500°, 4000° and 4500°ᴋ. The spectral-energy distribution tables of Skogland (1929) and Frehafer and Snow (1926) were used as sources of values of $J_\lambda/J_{0.59}$ at a series of wave-lengths in the visible region Skogland's tables, using the value $C_2 = 14330$ micron degrees, cover the range 2000°ᴋ. to 3120°ᴋ. in steps of 20°; they have been used for temperatures up to 3120°ᴋ., with the corresponding value of C_2. Frehafer and Snow's tables, using the value $C_2 = 14350$ micron degrees, cover the range 1000° to 28000°ᴋ.; they have been used, with the corresponding value of C_2, for temperatures above 3120°ᴋ.; their interval is 250° in the relevant range. To make the two sets of data accurately comparable, all the temperatures used in connection with Skogland's tables might be multiplied by 14350/14320, but the corrections are at most six or seven degrees, and are smaller than the uncertainty in T_1; they have therefore been neglected.

The data obtained for the curves of $\left(\int_0^\infty J_\lambda dt \Big/ \int_0^\infty J_{0\cdot59}\, dt \right)_{T_0}$ for the several values of T_0 are given in table 1 below. For each value of T_0, there is given for comparison the data for curves of $(J_\lambda/J_{0\cdot59})_T$ against λ for suitable values of T, in the region for which the two spectral distributions are similar. Except for $T_0 = 4500°$ K., the middle value given for T is T_1, the value of T which gives the closest agreement between the spectral distribution of the time-integral of the energy-flux of a source cooling from T_0, and that of the energy-flux of a source

Table 1

(i)

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\, dt}{\int_0^\infty J_{0\cdot59}\, dt}$	$\dfrac{J_\lambda}{J_{0\cdot59}}$		
	$T_0 = 2500°$ K.	$T = 2180$	$T = 2200$	$T = 2220$
0·32	0·0029	0·0018	0·0019	0·0021
0·40	0·044	0·035	0·037	0·039
0·45	0·137	0·121	0·125	0·129
0·50	0·327	0·308	0·314	0·319
0·55	0·644	0·632	0·636	0·641
0·59	1·000	1·000	1·000	1·000
0·65	1·699	1·723	1·707	1·692
0·70	2·413	2·450	2·411	2·374
0·72	2·726	2·762	2·712	2·664
0·76	3·383	3·408	3·332	3·258

$T_0 = 2500°$ K.

(ii)

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\, dt}{\int_0^\infty J_{0\cdot59}\, dt}$	$\dfrac{J_\lambda}{J_{0\cdot59}}$		
	T_0 $3000°$ K.	$T = 2540$	$T = 2560$	$T = 2580$
0·32	0·011	0·0067	0·0071	0·0076
0·40	0·094	0·074	0·077	0·080
0·45	0·225	0·198	0·202	0·207
0·50	0·435	0·409	0·415	0·420
0·55	0·723	0·709	0·712	0·716
0·59	1·000	1·000	1·000	1·000
0·65	1·470	1·490	1·479	1·469
0·70	1·884	1·912	1·890	1·868
0·72	2·053	2·078	2·050	2·023
0·76	2·384	2·395	2·356	2·317

$T_0 = 3000°$ K.

(iii)

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\, dt}{\int_0^\infty J_{0\cdot59}\, dt}$	$\dfrac{J_\lambda}{J_{0\cdot59}}$		
	$T_0 = 3500°$ K.	$T = 2840$	$T = 2900$	$T = 2960$
0·32	—	0·016	0·018	0·021
0·40	0·122	0·120	0·134	0·142
0·45	0·318	0·271	0·286	0·302
0·50	0·531	0·491	0·507	0·522
0·55	0·783	0·763	0·773	0·782
0·59	1·000	1·000	1·000	1·000
0·65	1·323	1·357	1·335	1·315
0·70	1·586	1·632	1·587	1·546
0·72	1·684	1·732	1·678	1·627
0·76	—	1·912	1·832	1·770

$T_0 = 3500°$ K.

(iv)

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\, dt}{\int_0^\infty J_{0\cdot59}\, dt}$	$\dfrac{J}{J_{0\cdot59}}$		
	$T_0 = 4000°$ K.	$T = 3120$	$T = 3250$	$T = 3500$
0·32	—	—	—	—
0·40	0·190	0·173	0·200	0·257
0·45	0·412	0·344	0·378	0·446
0·50	0·615	0·563	0·595	0·654
0·55	0·831	0·806	0·824	0·856
0·59	1·000	1·000	1·000	1·000
0·65	1·230	1·265	1·230	1·171
0·70	1·394	1·447	1·381	1·270
0·72	1·454	1·509	1·429	1·298
0·76	—	—	—	—

$T_0 = 4000°$ K.

(v)

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\,dt}{\int_0^\infty J_{0\cdot 59}\,dt}$	$\dfrac{J_\lambda}{J_{0\cdot 59}}$		
	$T_0=$ 4500° K.	$T=$ 3250	$T=$ 3500	$T=$ 3750
0·32	—	—	—	—
0·40	0·245	0·200	0·257	0·320
0·45	0·504	0·378	0·446	0·514
0·50	0·692	0·595	0·654	0·712
0·55	0·871	0·824	0·856	0·886
0·59	1·000	1·000	1·000	1·000
0·65	1·161	1·230	1·171	1·123
0·70	1·266	1·381	1·270	1·182
0·72	1·302	1·429	1·298	1·196
0·76	—	—	—	—

(with header $T_0=4500°$ K. spanning the table)

at constant temperature T, over the visible range. The other two are values, on either side of T_1, for which the agreement definitely appears inferior; they give some measure of the uncertainty in T_1.

It will be seen that, for each temperature T_0, some temperature T_1 can be found such that the two spectral distributions agree fairly closely. The discrepancies (with the best values of T_1) increase with T_0, and so does the uncertainty in T_1. Thus, when $T_0=2500°$ K. the maximum discrepancy (over the wave-length range considered) is about 1 % of the value at $0\cdot59\,\mu$, and T_1 can be determined within $\pm20°$; but when $T_0=4500°$ K., the maximum discrepancy is about 4 % of the value at $0\cdot59\,\mu$, and T_1 can only be fixed within about $\pm150°$ (agreement at the lower wave-lengths is best with $T_1=3750°$ K., at the higher wave-lengths with $T_1=3500°$ K.). It is, however, evident that, at least up to $T_0=4000°$ K. the spectral distribution of the time-integral of the energy flux will agree within the usual experimental error with the spectral distribution of the energy-flux for some temperature T_1. The values of T_1 corresponding to various values of T_0, and the approximate limits of the uncertainties in T_1, are summarized in the following table:—

Table 2

T_0 (° K.)	T_1 (° K.)	T_1 calc. from (9) ($450+0\cdot70\ T_0$)
2500	2200± 20	2200
3000	2560± 20	2550
3500	2900± 50	2900
4000	3250±100	3250
4500	3600±150	3600

These values of T_0 and T_1 obey closely the following linear relation:

$$T_1 = 450 + 0.70 T_0, \qquad \dots\dots(9)$$

as may be seen from the third column in table 2. Over the range $T_0 = 2500°$ K. to $4500°$ K., therefore, the value of T_0 can be determined from the value of T_1, corresponding to the experimentally-determined spectral distribution of the time-integral of the energy-flux, by using equation (9).

If the radiation from the sources mentioned in the Introduction is in fact Planckian within small limits, and if the other assumptions of the treatment are justified, the values cited for the "temperature" should be increased by amounts between $500°$ K. and $1000°$ K. to obtain the value of the maximum initial temprature T_0. The temperature so obtained for the photoflash lamp is above the boiling point of aluminium oxide, which is in the region of $3000°$ K. (Ruff and Konschak, 1926). This suggests that the radiation is possibly not Planckian, and that the emissivity decreases with increase of wave-length in the visible range. This would be compatible with the work of Liebmann (1930) on the emissivity of aluminium oxide. Current experimental work on emissivities may throw further light on the matter. Further application of the theory must await the completion of this work and of various studies on the burning of aluminium in oxygen.

REFERENCES

ARENS, H. and EGGERT, J., 1930. *Z. wiss. Photogr.*, 28, 169.
CALDIN, E. F., 1946 a. *Proc. Phys. Soc.*, 58, 207.
CALDIN, E. F., 1946 b. *Proc. Phys. Soc.*, 58, 341.
DZIOBEK, Z. F., 1928. *Z. wiss. Photogr.*, 25, 287.
EDER, J. M., 1930. *Z. wiss. Photogr.*, 27, 337.
FORSYTHE and EASLEY, 1934. *Phys. Rev.*, 45, 123.
FREHAFER and SNOW, 1926. *U.S. Bur. Misc. Pub.* No. 56.
LIEBMANN, G., 1930. *Z. Phys.*, 63, 404.
O'BRIEN and RUSSELL, 1933. *J. Opt. Soc. Amer.*, 23, 118.
RUFF and KONSCHAK, 1926. *Z. Elektr.*, 32, 515.
SKOGLAND, J. F., 1929. *U.S. Bur. St. Misc. Pub.* No. 86.

DISCUSSION

on papers by E. F. CALDIN entitled:—

" The relation between the brightness and temperature of a total radiator." *Proc. Phys. Soc.*, 57, 440 (1945).
" The ratio of the new international lumen to the lightwatt." *Ibid.*, 58, 207 (1946).
" Light emission during cooling of a Planckian radiator." *Ibid.*, 58, 341 (1946).
" The determination of the initial temperature of a cooling total radiator from measurements of the spectral distribution of the energy emitted during cooling." *Ibid.*, 58, 350 (1946).

Dr. A. G. GAYDON. I should like to enquire if Mr. Caldin has studied the spectrum of his source in detail. Burning magnesium ribbon emits strong MgO bands, and I should have expected emission of the green system of AlO in the burning of aluminium foil in oxygen. This might affect Mr. Caldin's determinations of the temperature.

Mr. L. T. MINCHIN. I should like to ask two questions : (a) What is a lightwatt ? (b) Why is the possibility of chemiluminescence excluded from Mr. Caldin's discussion of the Mg flash ?

Even if the spectrum is continuous, might not the chemiluminescence spectrum be superimposed on incandescence ? Since the body is cooling rapidly whilst the spectral distribution of energy is being recorded, it must be difficult to be certain that it conforms entirely with temperature radiation.

Dr. L. HARTSHORN. I fail to see any excuse for the term " lightwatt ", and hope that physicists will refuse to adopt it. Although the author has disclaimed responsibility for its introduction, by using it he is helping to burden physics with an unnecessary unit. I gather that the quantity concerned is power weighted according to its effect on the eye. However it is weighted, it remains power, and it is difficult to see why the universally accepted unit of power, the watt, is not sufficient. Perhaps some qualifying term like " luminous power " is needed for the quantity, but it is surely out of place in the unit.

Prof. A. O. RANKINE. I agree with others that it is a mistake to use the term " lightwatt " as an apparently new unit, although there is no difference dimensionally from the general unit of power. We do not, and, I hope, will not, speak of " electrical watts ", " mechanical watts ", or " thermal watts ", for which there would also be no justification.

AUTHOR'S reply. (a) *Mechanism of light emission.* The purpose of the third and fourth papers under discussion is to derive theoretically the intensity-time curves and spectral distributions of energy to be expected from cooling Planckian radiators. Until experimental investigations now in progress in various laboratories are completed, one cannot adequately compare this theoretical model with experiment. Therefore I do not wish to rule out chemiluminescence as a possible mechanism. The spectroscopic studies of O'Brien and Russell (1933), Forsyth and Easley (1934), and van Liempt and de Vriend (1934, 1935) did not reveal any important contribution to the total energy from MgO or AlO bands.

(b) *Photometric units.* The term " lightwatt " was used by Ives in 1917, and given wider currency by Moon in 1936. If we consider an area A cm.^{2} traversed by a uniform radiant flux, the total energy flux is given by $A\int_0^\infty J_\lambda d\lambda$, where $J_\lambda d\lambda$ is the radiant flux per cm.^{2} in the wave-length range $\lambda\pm\frac{d\lambda}{2}$. The luminous flux, according to the international definition (*International Lighting Vocabulary*, C.I.E., 1938) is then $A\int_0^\infty V_\lambda J_\lambda d\lambda$, where V_λ is the value of the relative luminosity factor for the human eye at wave-length λ. (Representative values for V_λ have been internationally accepted since 1924.) If $J_\lambda d\lambda$ is expressed in watts cm.$^{-2}$, the above expression is said to give the luminous flux in lightwatts. The term is thus not analogous to the term " mechanical watts ", to which objection is rightly taken ; for the ratio of $\int_0^\infty V_\lambda J_\lambda d\lambda$ to $\int_0^\infty J_\lambda d\lambda$ depends on the spectral distribution of J_λ, and so is not constant like the ratios of mechanical, electrical, and thermal units of power. The main objection to the use of the term is that the values of V_λ have no special physical status ; they are related to the response of a particular type of receiver under particular conditions, and would be quite different if the receiver were, say, a photoelectric cell. The best specification is, no doubt, a curve or function expressing the spectral distribution of energy flux. But a special term for a frequently occurring integral, closely related to measurable quantities, is always a convenience, and in its absence one has to employ some clumsy device such as that used in the third of the papers under discussion, where intensities computed in terms of $\int_0^\infty V_\lambda J_\lambda d\lambda$ were all multiplied by a constant and expressed in candles. Such computations are of increased importance since the new international candle, defined in terms of the brightness of a total radiator at a fixed temperature, is expected to come into use fairly soon. Perhaps photometrists could suggest a more acceptable term.

REFERENCES

FORSYTH and EASLEY, 1934. *Phys. Rev.*, 45, 123.
IVES, H. E., 1917. *Astrophys. J.*, 45, 41.
MOON, 1936. *Scientific Basis of Illuminating Engineering* (McGraw-Hill).
O'BRIEN and RUSSELL, 1933. *J. Opt. Soc. Amer.*, 23, 118.
VAN LIEMPT and DE VRIEND, 1934. *Rec. Trav. Chim.*, 53, 839 ; 1935. *Ibid.*, 54, 239.

THE PROPAGATION OF SOUND BETWEEN WALLS OF POROUS MATERIAL

By R. A. SCOTT,

Manchester

ABSTRACT. Deficiencies of existing theories relating to the propagation of sound along ducts lined with porous sound-absorbing materials are discussed. Attention is drawn to the wave-nature of the propagation of acoustic disturbances in the lining material, and a new theory is put forward in which the propagation of acoustic waves in the composite system of air space and lining is investigated for two simple shapes of duct. The results of the theory are in agreement with those of an earlier theory by Morse as long as the attenuation in the duct is not large, but wide divergences appear, in particular where the attenuation in the duct begins to approach that characteristic of the propagation of sound through bulk lining material. In the latter case it is shown by a particular experimental example that the attenuation found in practice is in agreement with the theory described in this paper.

§ 1. INTRODUCTION

VARIOUS aspects of the problem of the transmission of sound through pipes and channels have been investigated in the past by theory and experiment. For pipes with smooth rigid walls, the primary effect of walls is to redirect part of the sound by reflection. Apart from the influence of viscosity on the thin layer of air near the wall, the energy of the waves is conserved during passage along the pipe and the sound is propagated with the velocity characteristic of unrestricted air. Pipes with non-rigid walls or with walls absorbent to sound can affect the free passage of sound waves considerably, and under appropriate circumstances the presence of the walls leads to rapid attenuation of the sound and to substantial modification of its velocity of propagation.

Theories of the absorption of sound waves in pipes lined with absorbing material have been formulated by several authors, with varying success, in terms of the acoustic impedance* of the lining material. For pipes or channels in which, for example, the attenuation is large, this approach is not valid, and in the work which follows, a more complete theory is provided for cases in which the walls are lined with porous material which is homogeneous and isotropic. This theory takes account of propagation of sound in the lining itself.

§ 2. PREVIOUS THEORIES AND THEIR LIMITATIONS

In the past, the behaviour of an absorbing material in a sound field has most frequently been described in terms of the absorption coefficient for normal incidence. This quantity is the fraction of the energy of a plane sound wave incident normally on the surface that is not reflected by the material. The absorption coefficient refers only to the transfer of energy from a wave which falls

---

* The use, as in this paper, of the description " acoustic impedance of the lining " to refer to the unit-area impedance at one surface of the lining when the other surface is backed by a substantially rigid wall, and when sound falls at normal incidence on the surface, is general in the American literature on this subject.

in a specified direction on a surface and, therefore, conveys no information about the treatment received by waves which are directed obliquely at the surface of the material, as in the case of waves passing through a pipe lined with the material.

In most recent theoretical studies in acoustics, as, for example, in the theories of Morse (1939 a) and others relating to the decay of sound in rooms, the absorption coefficient has been displaced in favour of the much more informative quantity— the acoustic impedance at the wall. This quantity is the complex ratio of the oscillatory pressure at the surface of the material to the resultant oscillatory velocity at the surface for a plane wave incident normally on the surface. In 1939 Sivian showed that by assuming that the oscillatory pressure across the pipe is uniform, the propagation characteristics of waves travelling down a pipe lined with absorbing material can be expressed in terms of the acoustic impedance of the lining and of the geometrical form and dimensions of the pipe. He compared the attenuation computed from his formula with the attenuation measured in a circular pipe lined with an inch-thick layer of rock-wool. Satisfactory agreement was obtained for low frequencies, but for higher frequencies (over 2000 c/s. for a pipe of radius 10 cm.) the theory failed to explain the comparatively low attenuation found in practice. The cause of this failure is not difficult to find, and lies in the assumption of uniform oscillatory pressure across the pipe. Rayleigh has shown that a general acoustic disturbance travels along a tube as a substantially plane wave only when the radius is less than 0·29 of the wave-length. For a tube with walls of sound-absorbing material, the pressure across the tube is unlikely to be uniform beyond frequencies for which the radius of the tube approaches the quarter wave-length of the sound, since at such frequencies diffraction effects are insufficient to allow uniform diffusion of the sound across the tube.

The main deficiency of Sivian's theory has been removed by Morse (1939 b), who has derived a theory of the attenuation of sound in uniform, rectangular and circular lined ducts for cases wherein the whole effect of the wall-lining on the wave can be described in terms of the acoustic impedance of the lining. Morse has derived expressions for the attenuation constant for each of the possible modes of propagation in the duct and has published comprehensive charts from which to evaluate the attenuation constant. Beranek (1940) has applied Morse's theory to certain cases of square and rectangular ducts lined with rock-wool and has found good agreement with experiment. Rogers (1940) has derived theoretical expressions which are substantially the same as those of Morse and, for the very special case of material possessing only a resistive component of acoustic impedance, has reduced the expressions to simpler terms.

Implicit in Morse's theory is the assumption that the motion at the surface depends only on the acoustic impedance and on the local acoustic pressure and not on the acoustic pressure elsewhere. For homogeneous material of the porous type, this assumption leads to results which frequently are not seriously in error. Nevertheless, the assumption is not strictly tenable, and it is proposed in the present paper to consider in more detail the behaviour of acoustic waves in a lined pipe.

It has been surmised (Crandall, 1927) that sound is propagated as an attenuated wave in a porous material, and recent work (Scott, 1946 a) by the author has demonstrated that this view leads to an accurate description of the acoustic disturbance in

the material.   Part of the initial energy of a wave which travels along a pipe lined with porous material is, therefore, transmitted through the lining itself.   The oscillatory velocity at the surface of the lining depends in consequence on the joint action of the wave disturbances inside and outside the lining, and cannot be expressed merely in terms of the acoustic impedance and the oscillatory pressure at the surface of the lining.   In the theory which is given below, the course of the wave is traced through the composite space of air passage and pipe lining combined. The resulting expressions show that, as in Morse's theory, the acoustic disturbance can be resolved into modes which possess fixed propagation constants.   Under certain conditions, notably when the attenuation in the duct is small and where the lining consists of a densely packed collection of fibres, the expressions for the propagation constant in the duct tend to the same form as do those of Morse. Where the attenuation is higher, as, for example, in narrow ducts and at high frequencies in the acoustic range, Morse's expression gives attenuations of much higher magnitude than are given by the more complete theory.

§3. THE PASSAGE OF ACOUSTIC WAVES ALONG A CHANNEL LINED WITH POROUS MATERIAL

In the theory that follows, the porous lining material is considered as homogeneous and isotropic in so far as concerns the propagation of acoustic waves. For such materials it has been shown, both on acceptable theoretical grounds (Monna, 1938, and Morse and Bolt, 1944) and by experiment (Scott, 1946 a), that any acoustic disturbance in the medium must satisfy a wave equation of simple type.

Thus if the vector $\mathbf{v}$, which will be called the velocity of average flow, represents the rate of volume flow through unit area of cross-section normal to $\mathbf{v}$ in the lining material, it can be shown (see Scott, *loc. cit.*) that the equation for the velocity potential of average flow ψ, defined so that

$$\mathbf{v} = -\operatorname{grad}\psi \qquad\qquad \ldots\ldots(1)$$

is

$$\nabla^2\psi + h^2\psi = 0; \qquad\qquad \ldots\ldots(2)$$

h is here a complex quantity and is simply related to the attenuation and phase velocity of plane sound waves in the porous medium.   Thus

$$jh = \alpha + j\beta,$$

where α is the attenuation constant and β is the wave-length constant of an acoustic wave in the bulk material.   The equation for the acoustic pressure in the pores of the porous material can be shown to be given by $p = \rho'\dot\psi$, and since the present concern is only with acoustic disturbances of fixed pulsatance ω, this may be written

$$p = j\omega\rho'\psi. \qquad\qquad \ldots\ldots(3)$$

ρ' is a second complex quantity which is related to the characteristic impedance of the lining material.   The parameters h and ρ' can be obtained experimentally from measurements performed on the material of the lining in bulk (Scott, *loc. cit.*).   Equations (1), (2) and (3) together describe the acoustical disturbance

in the lining of the channel. Corresponding expressions for the velocity potential of flow ϕ in the air passage are well known to be

$$\mathbf{v} = -\operatorname{grad}\phi, \qquad \ldots\ldots(4)$$

$$\nabla^2\phi + k^2\phi = 0, \qquad \ldots\ldots(5)$$

and
$$p = j\omega\rho_0\phi, \qquad \ldots\ldots(6)$$

where $\mathbf{v}$ is the oscillatory velocity in the air passage, while k is the wave-length constant $(2\pi/\text{wave-length})$ and ρ_0 is the density of air.

Determination of the nature of waves which travel along the channel involves the solution of equations (1) to (6) consistent with the boundary conditions imposed by the channel.
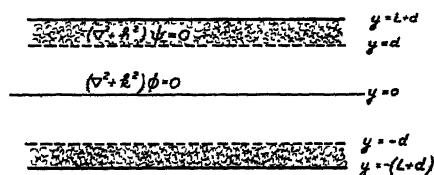
Figure 1. Plan of channel

Figure 1 represents a cross-section of the channel. Suppose the width of the air space is $2d$ and that the linings are each of depth L. Let x represent distances along the channel and y distances away from the central plane of the channel. The equation for ψ in the linings becomes

$$\frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial x^2} + h^2\psi = 0 \qquad \ldots\ldots(7)$$

and
$$p = j\omega\rho'\psi \qquad \ldots\ldots(8)$$

The equations for ϕ in the central air space become

$$\frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial x^2} + k^2\phi = 0 \qquad \ldots\ldots(9)$$

and
$$p = j\omega\rho_0\phi. \qquad \ldots\ldots(10)$$

For any one mode of propagation of a wave along the duct, ϕ and ψ will have a common *propagation constant* Γ such that

$$\phi = \phi(y)e^{-\Gamma x}e^{j\omega t}, \qquad \ldots\ldots(11)$$

$$\psi = \psi(y)e^{-\Gamma x}e^{j\omega t} \qquad \ldots\ldots(12)$$

Substituting these expressions in (7) and (9) and solving, we have:—

$$\phi(y) = Ae^{j\lambda y} + Be^{-j\lambda y} \text{ where } \lambda^2 = \Gamma^2 + k^2,$$
$$\psi(y) = Ce^{j\mu y} + De^{-j\mu y} \text{ where } \mu^2 = \Gamma^2 + h^2,$$

and where A, B, C and D are constants whose values are yet to be found.

When $y = 0$, $-\dfrac{\partial\phi}{\partial y} = 0$, by symmetry .

$$\therefore \quad A = B. \qquad \ldots\ldots(13)$$

At the outer wall, $y = L + d$ and the flow velocity normal to the wall $\left(-\dfrac{\partial\psi}{\partial y}\right)$ is zero;

$$\therefore \quad j\mu Ce^{j\mu(L+d)} - j\mu De^{-j\mu(L+d)}$$

or $$D = Ce^{2j\mu(L+d)} \qquad\qquad \ldots\ldots(14)$$

At the surface of the lining $(y = d)$, there must be continuity of the average flow velocity and of the oscillatory pressure, hence:

$$j\lambda Ae^{j\lambda d} - j\lambda Be^{-j\lambda d} = j\mu Ce^{j\mu d} - j\mu De^{-j\mu d}$$

and

$$\rho_0(Ae^{j\lambda d} + Be^{-j\lambda d}) = \rho'(Ce^{j\mu d} + De^{-j\mu d}).$$

Substituting for B and D from (13) and (14) and eliminating A and C,

$$\rho_0 \frac{\coth(-j\lambda d)}{\lambda} = \frac{\rho' \coth j\mu L}{\mu},$$

which may be written

$$\rho_0 \frac{\coth(-j\sqrt{\Gamma^2 + k^2}.d)}{\sqrt{\Gamma^2 + k^2}} = \frac{\rho' \coth j\sqrt{\Gamma^2 + h^2}.L}{\sqrt{\Gamma^2 + h^2}}. \qquad\qquad \ldots\ldots(15)$$

The equation (15), therefore, gives the propagation constant of the mode in terms of the dimensions of the duct and the parameters ρ' and h of the absorbing material.

§4. THE PASSAGE OF ACOUSTIC WAVES ALONG A CIRCULAR PIPE LINED WITH POROUS MATERIAL

Expressions can be obtained in a manner similar to that used above for the propagation constant associated with a mode of propagation along a circular pipe lined with a layer of uniform thickness of homogeneous, isotropic porous material. Figure 2 shows the cross-section of a pipe of radius b lined with a thickness $(b - a)$ of porous material. Equations (1), (2) and (3) again give the velocity potential of
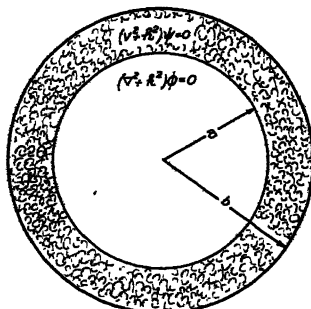
Figure 2.   Section of tube

average flow in the material of the lining, and (4), (5) and (6) give the velocity potential of flow in the air passage.   Let z represent distances along the axis of the pipe and let r represent radial distances from the axis.   Radial symmetry allows us to write

$$\nabla^2 = \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}\right).$$

If Γ again represents the propagation constant of any one mode, the velocity potentials may be written:

$$\phi = \phi(r)e^{-\Gamma z}e^{j\omega t}, \qquad \ldots\ldots(16)$$

$$\psi = \psi(r)e^{-\Gamma z}e^{j\omega t}, \qquad \ldots\ldots(17)$$

whence, by substitution in (2) and (5),

$$\frac{\partial^2\phi}{\partial r^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} + \lambda^2\phi = 0, \text{ where } \lambda^2 = \Gamma^2 + k^2,$$

$$\frac{\partial^2\psi}{\partial r^2} + \frac{1}{r}\frac{\partial\psi}{\partial r} + \mu^2\psi = 0, \text{ where } \mu^2 = \Gamma^2 + h^2.$$

The solutions for $\phi(r)$ and $\psi(r)$ are, therefore,

$$\phi(r) = AJ_0(\lambda r) + BY_0(\lambda r), \qquad \ldots\ldots(18)$$

$$\psi(r) = CJ_0(\mu r) + DY_0(\mu r). \qquad \ldots\ldots(19)$$

The boundary conditions are as follows:

$$r = 0, \quad \phi \text{ finite, whence } B = 0$$

$$r = a, \quad -\frac{\partial\phi}{\partial r} = -\frac{\partial\psi}{\partial r}, \quad \text{and}$$

$$j\omega\rho_0\phi = j\omega\rho'\psi$$

$$r = b, \quad -\frac{\partial\psi}{\partial r} = 0.$$

Hence, by differentiation, etc., of (18) and (19),

$$\lambda AJ_1(\lambda a) = \mu\{CJ_1(\mu a) + DY_1(\mu a)\} \qquad \ldots\ldots(20)$$

$$\rho_0 AJ_0(\lambda a) = \rho'\{CJ_0(\mu a) + DY_0(\mu a)\} \qquad \ldots\ldots(21)$$

and $\qquad 0 = \mu\{CJ_1(\mu b) + DY_1(\mu b)\},$

From the latter equation $D = -CJ_1(\mu b)/Y_1(\mu b)$, whence, by substitution in (20) and (21) and then eliminating A and C we find,

$$\frac{\rho_0}{\lambda}\frac{J_0(\lambda a)}{J_1(\lambda a)} = \frac{\rho'}{\mu}\frac{J_0(\mu a)Y_1(\mu b) - Y_0(\mu a)J_1(\mu b)}{J_1(\mu a)Y_1(\mu b) - Y_1(\mu a)J_1(\mu b)},$$

and since $\lambda^2 = \Gamma^2 + k^2$ and $\mu^2 = \Gamma^2 + h^2$, the equation for the propagation constant Γ becomes

$$\frac{\rho_0}{\sqrt{\Gamma^2+k^2}}\frac{J_0(\sqrt{\Gamma^2+k^2}\,.\,a)}{J_1(\sqrt{\Gamma^2+k^2}\,.\,a)} = \frac{\rho'}{\sqrt{\Gamma^2+h^2}}$$

$$\times \frac{J_0(\sqrt{\Gamma^2+h^2}\,.\,a)Y_1(\sqrt{\Gamma^2+h^2}\,.\,b) - Y_0(\sqrt{\Gamma^2+h^2}\,.\,a)J_1(\sqrt{\Gamma^2+h^2}\,.\,b)}{J_1(\sqrt{\Gamma^2+h^2}\,.\,a)Y_1(\sqrt{\Gamma^2+h^2}\,.\,b) - Y_1(\sqrt{\Gamma^2+h^2}\,.\,a)J_1(\sqrt{\Gamma^2+h^2}\,.\,b)}.$$

$$\ldots\ldots(22)$$

§5. DISCUSSION OF THEORETICAL RESULTS

Equations (15) and (22) give series of values of Γ which satisfy the boundary conditions for, respectively, a channel and a pipe lined with porous material. Each value of Γ is associated with a corresponding variation of ϕ and ψ across the channel

as given by (11) and (12) or by (16) and (17), and, therefore, refers to a particular "mode of propagation" of sound in the sytsem.   An arbitrary acoustic disturbance at the beginning of the channel or pipe can be resolved into a number of these component modes and the separate modes regarded as travelling along the tube, each with a wave-front of fixed shape, and with fixed attenuation constant and phase velocity.

Morse's theory relating to the propagation of sound in ducts leads to a similar multiplicity in values of Γ and, therefore, to the existence of many modes of propagation for a given duct.   Morse's theory, which would have held for the case of the channel lined with porous material if the motion at the surface of the lining had been always normal to the lining, and dependent solely on the oscillatory pressure near the lining, gives as the equation for Γ

$$\rho_0 \frac{\coth\left(-j\sqrt{\Gamma^2+k^2}.d\right)}{\sqrt{\Gamma^2+k^2}} = \frac{Z}{\omega}, \qquad \ldots\ldots(23)$$

where Z is the acoustic impedance of the lining.   It can readily be shown (Scott, 1946 b) that Z may be expressed in terms of ρ', h and L in the following manner :

$$Z = \frac{\rho'\omega}{h}\coth jhL. \qquad \ldots\ldots(24)$$

Morse's equation for Γ may therefore be rewritten in a form which compares easily with that obtained in this paper; thus (23) becomes

$$\rho_0 \frac{\coth\left(-j\sqrt{\Gamma^2+k^2}.d\right)}{\sqrt{\Gamma^2+k^2}} = \rho'\frac{\coth jhL}{h}. \qquad \ldots\ldots(25)$$

The two expressions (15) and (25) lead to the same values of the propagation constant Γ for cases in which Γ^2 is very small compared with h^2, that is, when the attenuation constant and the wave-length constant associated with the mode are much less than the attenuation constant and wave-length constant characteristic of the material of the lining in bulk.   Low values of attenuation constant are to be found in practice for wide air passages ($2d > 15$ cm. for linings of rock-wool).   For smaller values of d, Morse's formula gives attenuations which increase without limit as d is reduced.   Such high attenuations are not found in practice, nor are they to be expected, for it is clear that if the air-passage between two linings of porous material is reduced sufficiently, the passage of an acoustic disturbance along the channel is tantamount to the propagation of waves through a channel filled with the lining material.   In this case the value to be expected for Γ is that of the propagation constant jh of the porous material.   An experimental illustration of this point is given by the data of figure 3.   This figure shows measured values of attenuation (in decibels/cm.) determined for a channel with an air-space 1·27 cm. wide flanked by two 5·08-cm. thick linings of rock-wool (stillite of bulk density 0·08 gm./cm.³).   Curve A is the measured attenuation, B is the attenuation in the bulk material and C is the attenuation associated with the lowest mode of propagation calculated from Morse's formula.   The latter calculation is facilitated by the charts which are published in Morse's paper. The figure also shows the corresponding velocities of propagation of waves in the channel.   Thus E is the measured velocity, F is the velocity of a wave in the porous

material in bulk and G is the velocity calculated from Morse's formula for the fundamental mode. The experimental results were obtained under conditions in which the disturbance in the channel consisted substantially of the fundamental mode of propagation in the channel. The material used for the linings was closely similar to that used by the author in a recent paper upon the propagation of sound in porous media,* and the experimental values of the parameters h and ρ' therein contained are, therefore, relevant to this material.

The complete solution of equation (15) or of equation (22) involves a process of successive approximation, and in general is extremely tedious; only in exceptional cases is the considerable work of computation justified in practice. It is therefore important to investigate the conditions under which Morse's formula gives a value of Γ that for practical purposes is a sufficiently good approximation to the true value. The difference between the two expressions depends upon the difference between $(\Gamma^2 + h^2)^{\frac{1}{2}}$ and h, and for this difference to be small the
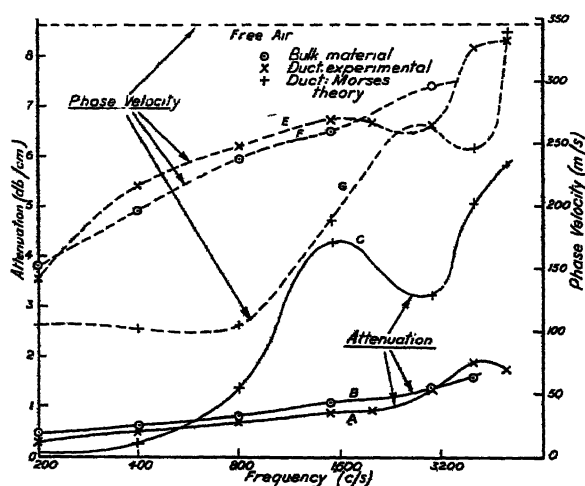
Figure 3. Comparison of results of experiment and Morse's theory for a narrow-lined duct.

real and imaginary components of Γ^2 must respectively be much less than the real and imaginary components of h^2. Thus if α_1 and β_1 represent the attenuation constant and wave-length constant corresponding to Γ, i.e. $\Gamma = \alpha_1 + j\beta_1$, and if, as before, $jh = \alpha + j\beta$,

$$\Gamma^2 = (\alpha_1{}^2 - \beta_1{}^2) + 2j\alpha_1\beta_1$$

and

$$h^2 = (\beta^2 - \alpha^2) - 2j\alpha\beta,$$

and, therefore, $(\alpha_1{}^2 - \beta_1{}^2)$ and $\alpha_1\beta_1$ must be respectively much less than $\beta^2 - \alpha^2$ and $\alpha\beta$. For most practical sound absorbents of the porous type, β and α are of the same order of magnitude at all frequencies, and in this case it is necessary that neither α_1 nor β_1 be comparable in magnitude to α or β. Thus, for Morse's expression to apply, the attenuation calculated therefrom must be substantially less than that characteristic of the bulk material, and the velocity of propagation of the sound (of a given mode) must be much higher than that characteristic of the bulk

* For the data relating to the material used, see tables 2 and 3 for h and ρ' and curves marked 2 in figure 6 for the acoustic impedance Z of the lining.

material.   Satisfaction of the first condition implies, in general, either that the quantity $Z/\rho c$ be large compared with unity, i.e. the impedance Z must be high compared with the characteristic impedance of air $(\rho_0 c)$, or that the duct width be large compared with the wave-length.   The existence of low attenuation in the duct does not, however, insure that the second condition be satisfied.   Where the ratio of the wave-length constant β of the bulk material to the wave-length constant k of unrestricted air is not very different from unity, β_1 is necessarily of the same order as β, and in consequence Morse's formula again fails.   This state of affairs is instanced by the results in the lower frequency range of figure 3.   In the case considered, the velocity of propagation in the bulk material is only about half that in free air, i.e. $\beta \simeq 2k$.   The value of velocity calculated from Morse's formula is lower than that in the bulk material, and in consequence this calculated value of β_1 is substantially greater than β, which is inadmissible if the results of Morse's formula are to be in agreement with those deduced from the more exact formula of the present paper.

It is instructive to consider the physical conditions which govern the sound field in the duct lined with porous material.   In Morse's theory it is assumed that the normal component of the acoustic velocity at the lining is given by the ratio of the oscillatory pressure at the surface to a fixed quantity (the acoustic impedance of the lining), irrespective of the intensity or direction of the wave in the duct. Whilst for a uniform wave falling obliquely on the surface of a densely packed material this relation is approximately true, it is not in general strictly tenable, and Morse's theory therefore contains the implicit assumption that there is no axial component of velocity at the surface of the lining.*   Moreover, for a sound wave propagated along a duct, the pressure at the surface is not uniform.   The attenuation of the sound in the duct implies that the acoustic pressure in the lining falls continuously in the direction of the axis of the duct, and in consequence there is a forward diffusion of sound in the lining.   The pressure gradient normal to the surface, and, therefore, the velocity component in this direction, is thereby dependent on the pressure at adjacent sections along the duct, and not solely on the local pressure as is required in Morse's theory.

The discrepancy involved in the application of Morse's theory to a duct lined with porous material can be related in a general way to the attenuation (α) and wave-length constant (β) associated with the propagation of uniform waves in the material.   The extent to which the pressure gradient normal to the surface of the lining is affected by the acoustic disturbance at preceding sections depends firstly on the direction of the oscillatory velocity in the lining.   In the air-space, the velocity vector is mostly longitudinal, but refraction of the wave at the lining tends to turn the vector towards the wall.   The " refractive index" at the boundary is clearly given by the ratio β/k of the wave-length constants in the two media.

---

* The author has shown by means of auxiliary experiments that the velocity in the air-space is closely axial even at the boundary surface.   For this purpose a shallow duct was built with lined walls contained between glass sheets, one-half inch apart, which served as top and bottom of the duct.   Sound was led into the duct at one end and the field, delineated by illuminated smoke-particles, was explored with the aid of a low-power microscope.   Over a wide range of width of air-space and duct-lining, and for all frequencies between 100 and 1000 c./s. it was estimated that the inclination of the particle traces to the axis did not exceed 5° at the boundary. See Morse and Bolt (1946), p. 92.

Where this ratio, and, therefore, the turning action are large, the pressure at a point immediately behind the lining surface is not as much affected by the wave at preceding sections as where the ratio is small. The gradient of pressure normal to the lining is, therefore, less dependent on the disturbance at adjacent sections in the former case than in the latter, and the application of Morse's theory is likely to be least in error where β/k is greatest.*

The disturbing effect on the normal gradient of pressure, due to waves falling on preceding sections of the duct, is clearly also dependent on the attenuation coefficient α of the lining material, since where α is high the magnitude of the disturbance at one section due to waves at an earlier section is correspondingly reduced. The application of Morse's theory is less likely to be in serious error where the density of packing of the material is great, and α is high, than where the material is loosely packed.

The above physical considerations illustrate in a qualitative manner the conditions under which Morse's analysis might be expected to be valid for ducts lined with porous material, and the conclusions are in satisfactory agreement with the deductions which have been made from the mathematical analysis of the present paper.

§6. CONCLUSIONS

The theory given above of the propagation of sound between walls lined with homogeneous, isotropic porous material provides expressions from which can be calculated the attenuation and the phase velocity of any of the possible modes of propagation. These modes are, by definition, acoustic disturbances in which the wave-front retains a fixed shape and distribution of intensity in its passage through the duct. The theory is based on the assumption that any acoustic disturbance in the porous lining can be described in terms of a wave equation analogous to that which is known to describe acoustic disturbances in air, but containing a complex parameter in place of the wave-length constant of the latter. The complex parameter implies that a plane wave in the porous material is propagated as an attenuated wave, and the real and imaginary components of this parameter are readily deducible from the experimentally determinable attenuation constant and wave-length constant characteristic at any given frequency of plane waves in the bulk material.

The problem of deducing expressions for the attenuation and wave-length constants of any given mode in the duct reduces to the simultaneous solution of the wave equations for the air-space and for the lining, subject to the appropriate boundary conditions dictated by the geometry of the duct. Solutions to this problem are provided for the cases of a channel and a circular pipe lined symmetrically with porous material. Numerically these expressions can be solved in the general case only with difficulty. A number of important deductions are, however, readily made. Thus it is shown in the paper that the expression for the propagation constant is, under special conditions, substantially in agreement with that

* For commercially available sound-absorbent materials, β/k is rarely very large, except for very dense materials and at frequencies low in the acoustic range. Thus experiments show that for very densely packed rock-wool, the phase velocity in the material is about 40 m/s. at 100 c./s., which implies a refractive index of 8·6 at the boundary surface. The refractive index for commoner materials usually ranges from 1 at high frequencies to 3 at low frequencies.

deduced earlier by Morse on the assumption that the effect of the linings on the passage of sound through the duct can be fully described in terms of the acoustic impedance of the lining. For ducts lined with densely packed sound-absorbing material, and for which the duct width is not small compared with the wave-length of sound in free air, Morse's formulae give results which are in reasonable agreement with those deducible from the formulae of the present paper. Numerical solutions of Morse's formulae can be obtained without undue difficulty from the comprehensive set of charts provided in Morse's paper on this subject (*loc. cit.*).

Where the sound-absorbing lining consists of a loosely packed material, or where the width of the air-space in the duct is small compared with the wave-length, Morse's formulae no longer give a reasonably good approximation to the proper solution of the problem. Thus, in particular, the attenuation constant calculated from Morse's formulae is infinite for the case of a vanishingly small air-space, whereas it is clear from general physical considerations, as also from the expressions deduced in the present paper, that in this case the acoustic disturbance is propagated with the attenuation constant characteristic of the propagation of sound in the porous material of the lining. Experimental results are included in the paper to illustrate this point.

§7. ACKNOWLEDGMENT

REFERENCES

BERANEK, L. L., 1940. *J. Accust. Soc. Amer.*, 11, 228.
CRANDALL, I. B., 1927. *Vibrating Systems and Sound*, p. 197 (London: Macmillan).
MONNA, A., 1938. *Physica*, 5, 129.
MORSE, P. M., 1939 a. *J. Accust. Soc. Amer.*, 11, 56.
MORSE, P. M., 1939 b. *J. Accust. Soc. Amer.*, 11, 205.
MORSE, P. M. and BOLT, R. H., 1944. *Rev. Mod. Phys.*, 16, 69.
ROGERS, R., 1940. *J. Accust. Soc. Amer.*, 11, 480.
SCOTT, R. A., 1946 a. *Proc. Phys. Soc.*, 58, 165; 1946 b. *Ibid.*, 58, 253.
SIVIAN, L. J., 1939. *J. Accust. Soc. Amer.*, 11, 56.

CATHODO-LUMINESCENCE: PART I. GROWTH AND DECAY PROCESSES

By J. W. STRANGE and S. T. HENDERSON,

Electric and Musical Industries, Ltd., Hayes, Middlesex

ABSTRACT. The active centre theory of luminescence, with certain assumptions, leads to the expectation of exponential growth and decay processes in light emission. Measurement has shown the widespread occurrence of exponential processes, though the details are much more complex than the simple theory would suggest. The effects of changing current density and temperature were examined. The phosphors studied fell into two classes, distinguished mainly by the rates of the processes measured. Some non-exponential processes were found, especially in long time decay, but none were definitely bimolecular.

§ 1. INTRODUCTION

THE work described in this and the following paper was mostly done in 1937–1940. References to parts of it have been published elsewhere (Strange, 1938, 1940). It is concerned with the phenomena of light emission in luminescent materials under cathode-ray bombardment, especially with the very rapid changes which occur during this process and with the changing efficiencies observed as conditions are varied. Previous work has largely concentrated on excitation by ultra-violet radiation, and slow phosphorescent decays have been studied far more than the more rapid processes.

It is desirable to define some terms that will be used, though the outline of the *active centre* theory of luminescence will be assumed already known. A *phosphor* or luminescent material means here a solid inorganic base material which has been heat-treated, generally with a small addition of some foreign metal or its salt (the *activator*) in such a way that the product emits visible light under a beam of cathode rays, ultra-violet or other radiation. The mechanism is a process of raising the atoms of the phosphor to higher energy states. Their return with the emission of light (*luminescence*) may be classified as *fluorescence* or *phosphorescence* according to the characteristics observed, the former being comparatively rapid and due to electrons making inner transitions, atomic or molecular, and the latter being temperature dependent owing to the metastable states involved (Pringsheim, 1939). The word *excitation* must be restricted to the process by which the impinging radiation raises electrons from ground states to higher levels of energy. The light emission at any moment depends on the number of electrons in excited states and on the probability of their return to ground states: the emission increases from zero when excitation begins, but attains its final value only after a finite time. This is the *growth process*. When light output has thus become constant under steady excitation, the *equilibrium state* has been reached. After the exciting radiation has been cut off, the light emission declines to a low value in a time generally comparable with that taken in the growth process; this is called the *decay process*.

We may refer briefly to Becquerel's work (1861). He demonstrated the rapid exponential decay with time of the light emitted by uranyl salts (fluorescence), while the slow phosphorescent decay of zinc sulphides gave the relation

$$cI_0{}^m = I^m(t+c), \qquad \ldots\ldots(1)$$

where c and m are constants, I_0 and I the intensity of light emitted respectively at the start and after a decay time t. If $m = \frac{1}{2}$, the usual bimolecular form results:

$$I_0 = I(1+\gamma t)^2. \qquad \ldots\ldots(2)$$

Lenard (1928) attempted to explain all long-period decays as due to the sum of several exponential processes with different constants. This is an unsatisfactory way of interpreting an experimental curve, and to a limited extent the same difficulty occurred in the present work.

Among the few modern papers giving results of cathodo-excitation of phosphors are those of Schleede and Bartels (1938), and Nelson, Johnson and Nottingham (1939). De Groot's work (1939), though concerned with ultraviolet excitation, is of considerable interest.

From the active-centre theory may be derived expressions for light output during and after the excitation period, if certain additional assumptions are made. The treatment given is due to Dr. L. F. Broadway, and it was the original motive for the experimental work described below. A similar theoretical deduction has since been published by Nelson, Johnson and Nottingham (1939). No use has been made of recent work on trapping states, published since this paper was written (Randall, Wilkins, Garlick, 1945), but this new work does not appear to afford so direct an approach to the problem as the discussion which follows. See also the introduction to Part III of the present series.

§2. THEORETICAL

(a) *Monomolecular or random type of process*

Consider the excitation of a thin slab of the phosphor, of unit area and thickness δy, at a depth y from the surface, and having n active centres per unit volume, of which number x are in excited states at time t from the start of excitation. The slab is assumed to be under a constant beam of i electrons per unit area per unit time, and the incident voltage of these electrons, V, is assumed constant (i.e., the straggling and scattering effects when $y > 0$ are neglected). Two probability constants are introduced: A for the excitation of centres by incident electrons at voltage V, and B for the return of excited electrons to lower states with emission of radiation. A must vary with y, but is here assumed constant, since δy is very small. A and B are both taken to be independent of i.

Then in time δt the number of active centres excited in unit volume is

$$\delta_1 x = Ai(n-x)\delta t.$$

The number of centres decaying with radiation in the same time is $\delta_2 x = Bx . \delta t$. Hence the net number of centres excited is

$$\delta_1 x - \delta_2 x = \{Ai(n-x) - Bx\}\delta t,$$

and the "growth process" is given by

$$\frac{dx}{dt} = Ain - x(Ai - B)$$

or $\quad x = Ain\{1 - \exp[-(Ai+B)t]\}/(Ai+B).$ ......(3)

Thus at time t the light output in quanta/unit vol./unit time from the thin slab of phosphor is

$$L = Bx = ABin\{1 - \exp[-(Ai+B)t]\}/(Ai+B), \qquad ......(4)$$

which requires an exponential rise of L as the excitation continues, with a *growth constant* $Ai + B$. At the end of excitation the constant B is alone concerned, and an exponential decay occurs, with *decay constant* B.

The total light output from a slab of finite thickness, or $\int L \cdot dy$, neglecting the absorption and scattering of light and the possible decrease of i as electrons penetrate the slab, will be assumed to be of the same form as equation (3) for the experimental conditions used, namely, a constant incident voltage at $y = 0$ and, complete absorption of the beam, which implies that the value of A applicable to the whole slab of phosphor bears a simple relation to A used above. The new value of A is denoted by A'.

At equilibrium excitation, where t is large, the light output becomes

$$L = A'Bin/(A'i + B), \qquad ......(5)$$

from which it appears that L is proportional to the current density when this is small, while at very high values of i the light output reaches a constant value Bn.

(b) *Bimolecular type*

If the active centre is ionized, the rate of recombination or decay, and hence the light output, will be proportional to Bx^2 instead of to Bx, and by an argument similar to the above, the light output L at equilibrium excitation is given by

$$L(2B) = 2A'Bin + (A'i)^2 - A'i\sqrt{\{(A'i)^2 + 4A'Bin\}}. \qquad ......(6)$$

Again L is proportional to the current density when this is small. If there are x_0 excited centres at the end of excitation, it is readily shown that, at low current densities,

$$x_0 = \sqrt{(A'in/B)}.$$

With the usual symbols,

$$-\frac{dx}{dt} = Bx^2 = I.$$

Hence $I_0 = I(Bx_0t + 1)^2$, and from equation (2) $\gamma = \sqrt{(A'B)}$ at low current densities, when it also follows that the growth curve is at first convex towards the time axis.

§3. EXPERIMENTAL METHODS AND RESULTS

(a) *Processes lasting up to 1 millisecond*

The apparatus consisted of a master oscillator generating frequencies from 500 to 2500 c.p.s., which controlled a pulse generator giving a train of rectangular pulses of about 100 v. amplitude over the same frequency range. These pulses applied to the modulator of the electron gun in the cathode-ray tube switched the electron beam on and off at the frequency required, giving pulses of light from the

luminescent screen.   A final anode voltage of 5 kv. was used in all the work to be described, except where otherwise stated.

The on and off periods were originally equal, but later the off period was
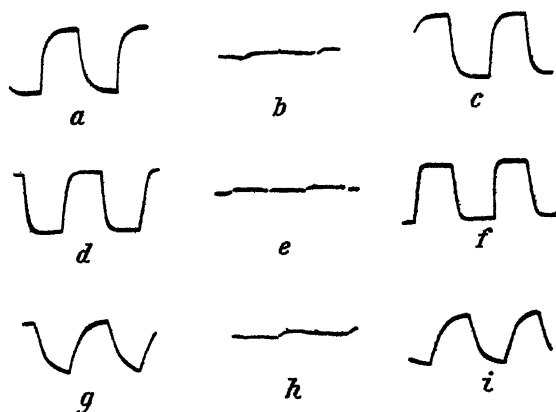
Figure 1.   Null method of estimating growth and decay constants.   Abscissae represent time ; growth or decay period = 500 μ sec.

<div align="center">

a. Photocell output, ZnS . Ag, 12 μa./cm²

b.   „   „   cancelled.

c. Cancelling signal.

Similarly *d, e, f* for ZnS . Ag, 180 μa./cm²

g, h, i for CaO . Sm, 10 μa./cm²

</div>
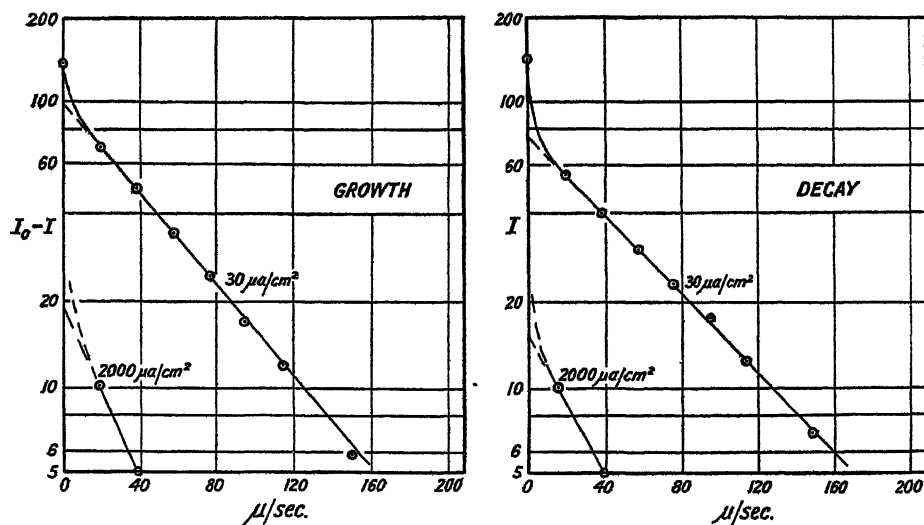
Figure 2.   Growth and decay for ZnS . Ag : semi-logarithmic plots.   Light units are arbitrary and different for different current densities.

increased to lessen the effect of incomplete decay of the phosphor sometimes observed.

The light emitted was received by a photocell and amplifier, and the amplified signals applied to one pair of plates in a 3-inch electrostatic cathode-ray monitor tube. Both parts of this recording end of the apparatus had an overall frequency

response which was flat up to one megacycle per second, a necessary condition to avoid "aperture" effects. The cathode-ray tube with the test samples was of demountable construction to provide for rapid changes of phosphors during operation and of the electron gun when required.

The apparatus described represented the growth and decay curves on a linear time base as shown in figure 1, *a, d, g*. These were photographed at different current densities and worked out graphically to give growth and decay curves such as figure 2 for ZnS.Ag. The fundamental exponential character is clear.

Later improvements extended the range of working down to currents of $1 \mu a./cm^2$, and a null method was also developed which allowed the feeding-in of one or more exponential voltage signals, derived by charge or discharge of condensers, to the recording monitor, so as to cancel the original trace; growth and decay could be separately treated if required, and the exponential time constants thus determined much more readily than by the photographic methods, and at least as accurately. Perfect cancellation was not usually obtained, but figure 1 shows some typical results. The left-hand column gives the cathode-ray tube trace produced by the photocell output for the material and current density indicated, the right-hand column the corresponding cancelling signals as recorded without the photocell output, and the centre column gives combinations of the two.

General survey of Ag class phosphors

It appeared at once that the lowest available frequency was too fast to give any results from manganese activated materials (see § 3 (b) below), while many oxides, sulphides, tungstates and silicates with other activators (or none) gave similar results to ZnS.Ag with the present apparatus, and the term *Ag class* was adopted to include all such phosphors. At the lowest current densities, the brightest of these gave non-exponential processes, possibly bimolecular. In the more easily accessible range, exponential growth and decay of similar time constants were found quite generally, with a faster initial exponential process in each (figure 1), which increased in relative proportion with increase of current density. The reservation already made regarding the validity of sums of exponentials must be borne in mind in examining table 1, which gives a general survey of decay constants found at a current density of $5 \mu a./cm^2$ (except for CaO.Sm at $10 \mu a./cm^2$). The fast and slow processes are named α and β respectively. Complex emission spectra (bands and fairly sharp lines) of the samarium-activated phosphors were here treated as a whole, but for materials described in section 3 (b) it was sometimes possible to separate out the decay processes for different parts of the emission spectrum.

ZnS.Ag, the material most used in this work, and typical of the class, contained 0·01% Ag and was activated with a chloride flux at about 900° c.

The noteworthy features of these results are: the very rapid decay of zinc oxide; the similarity of the tungstates and, in a different range, of the zinc sulphides; and the independence of these measured processes with respect to long-time phosphorescence, such as that induced by lead in tungstates, or by copper in zinc sulphides. Further, it was shown that ZnS.Ag containing up to one part in a million of nickel suffered no appreciable changes in these decay constants,

although the phosphorescence was drastically reduced by the nickel. This supports the view that while both ionization and random (i.e. exponential)

Table 1. Decay constants and proportions of processes

Material	Constant of α process (sec.$^{-1}$)	Constant of β process (sec.$^{-1}$)	% α	% β
ZnO	$1 . 10^6$	—	95	5
ZnSe	$1 . 10^6$	—	90	10
$CaWO_4$. Sm	$1 . 10^6$	$7 . 10^4$	5	95
$CaWO_4$	$1 . 10^6$	$6 . 10^4$	40	60
$CdWO_4$	$1 . 10^6$	$6 . 10^4$	40	60
$ZnWO_4$	$1 . 10^6$	$6 . 10^4$	20	80
$ZnWO_4$. Pb	$1 . 10^6$	$6 . 10^4$	20	80
SrS . Bi	$2 . 10^5$	$3 \cdot 5 . 10^4$	60	40
SrS . Ag	$1 . 10^5$	$1 \cdot 4 . 10^4$	60	40
CaS . Cu	$1 . 10^6$	$2 . 10^4$	70	30
ZnS . CdS . Cu (45 % Zn . S)	$1 . 10^5$	$1 \cdot 5 . 10^4$	30	70
ZnS .	$1 . 10^5$	$1 \cdot 3 . 10^4$	50	50
ZnS . Cu	$1 . 10^5$	$1 \cdot 3 . 10^4$	50	50
ZnS . Ag	$1 . 10^5$	$1 \cdot 3 . 10^4$	50	50
ZnS . Pb	$1 . 10^5$	$1 \cdot 3 . 10^4$	50	50
ZnS . Au	$1 . 10^5$	$1 \cdot 3 . 10^4$	50	50
Zn_2SiO_4	$7 . 10^4$	$1 \cdot 0 . 10^4$	40	60
Zn_2SiO_4 . Ag	$7 . 10^4$	$1 \cdot 0 . 10^4$	40	60
CaO . Sm	—	$9 \cdot 5 . 10^3$	5	95

processes may occur in these materials, they are acting in parallel and not consecutively.

Temperature effects

Other evidence on the nature of the processes was provided by measurements at low and high temperatures. For low temperatures, small sealed-off cathode-ray tubes were tested at 20° c. and again when the screen end was immersed in liquid oxygen. For both ZnS . Ag and $CdWO_4$ at 100 μa./cm² the α and β decay constants were unchanged by cooling, but a small increase in the proportion of the faster process occurred for each at $-180°$c. Observations above room temperature were made on a demountable tube with the screen end heated in a tubular oven Up to 150° c. ZnS . Ag showed a gradual increase of growth and decay " constants " up to twice their original values, and ZnS . Cu gave a steady decay constant up to 150° c., with approximate doubling of its value between 150° c. and 250° c. It is doubtful if the variations are really significant in view of the widely changing efficiencies of the phosphors with temperature, and the much greater variation with current density described below. We may say that the rates of growth and decay are not profoundly altered by temperature changes, in agreement with the view that random electron transitions are responsible for these processes.

Current density effects

Apart from changes in proportions of fast and slow decay processes, indication were soon met that current density changes also affected the values of the " con

stants" of growth and decay. Table 2 gives decay figures for ZnS.Ag. The growth figures were similar in trend, with somewhat greater values for the constants.

Table 2. Decay of ZnS.Ag

Constants of processes			Percentages of processes	
$\mu a./cm^2$	α (sec.$^{-1}$)	β (sec.$^{-1}$)	α	β
5	1.10^5	$1\cdot3 \;.\; 10^4$	50	50
12	1.10^5	$1\cdot45\;.\;10^4$	63	37
60	1.10^5	$1\cdot65\;.\;10^4$	75	25
180	$1\cdot2\;.\;10^5$	$4\cdot1\;.\;10^4$	87	13
700	$1\cdot8\;.\;10^5$	$5\cdot0\;.\;10^4$	90	10
5000	$4\cdot6\;.\;10^5$	$1\cdot0\;.\;10^5$	95	5

From the above results we are justified in assuming the widespread occurrence of exponential processes in this class of phosphors, but it is clear that the simple theory is far from accounting for the details. The other class of phosphors, the *Mn class*, will now be considered.

(b) *Processes lasting from 1 millisecond to 1 second*

Another circuit, based on a similar idea to that employed in the measurement of the constants of the Ag class, was used for this part of the work. The general layout of the apparatus is shown in figure 3; and a description is in the Appendix (p. 382).
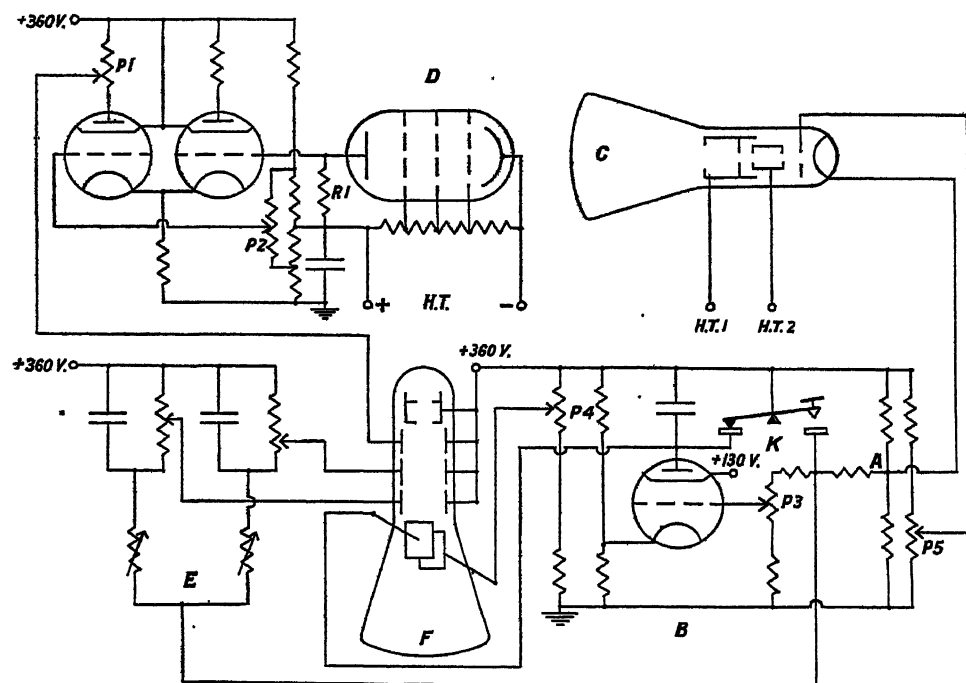
Figure 3. Luminescence decay apparatus.

The depression of the key K started simultaneously the following four processes :—(a) the scanning of the monitor tube, (b) the decay of the screen under test after equilibrium excitation, (c) the rise of potential in the output stage of the photoelectric unit due to decrease in the luminescent light, and (d) the exponential fall in potential across the condensers of the matching unit, balanced against the rise in (c).

The operation of the apparatus consisted of adjusting the current in the tube to the required value, interposing filters between the screen and the photocell unit to bring the output signal within the linear range of the amplifier, and then matching this signal with one or more signals from the matching unit, varying them in amplitude and constants until an accurate null value was obtained.
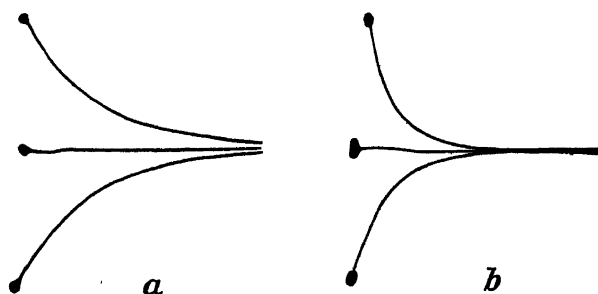
Figure 4.　Growth process for $Zn_2SiO_4 . Mn$.　Abscissae represent time ;　total period shown $\simeq$ 40 m.sec.　(a) 32 $\mu a./cm^2$;　(b) 200 $\mu a./cm^2$　Photocell output below, signal from matching unit above, mixture central in each case.

The circuit given is for the measurement of decay, but simple alterations can be made to adapt it to the measurement of growth.

One of the main difficulties in the operation of the apparatus was the accurate mixing of signals of large amplitude.　A satisfactory solution was obtained by using a monitor tube with four sets of plates; the three pairs used for the photocell output and the matching signals were parallel, and the fourth pair, used for scanning, was at right angles.　One plate of each pair was connected to the second anode of the tube.　The most important factor limiting accuracy was deflexion defocusing, due to voltage differences between the plates.　Examples of the traces obtained for growth of $Zn_2 SiO_4 . Mn$, using this null method, are reproduced in figure 4.

General survey of Mn class phosphors

By using no matching signals, the curve of growth or decay could be obtained on the monitor, photographed and measured.　Decay curves for $Zn_2 SiO_4 . Mn$ agreed closely with exponentials, apart from a small proportion of an initial faster process, as in the Ag class (see figure 5).

Manganese-activated materials showed the variations from simple behaviour already noted for the Ag class of phosphors; their decay constants increased slowly with current density, as did the proportion of faster decay process, which was also apparently exponential.　Zinc aluminate with manganese activator was peculiar in having very little fast decay, even at high current densities, and later work on $Al_2O_3 . Cr$ (see opposite page) afforded a similar case.

The willemite used was Zn_2SiO_4 with 0·3% Mn. The ZnS.Mn contained 0·1% Mn and was activated at about 1100° c.

The nature of the growth process differed from the simple theory even more than did the decay process. The growth constant was about the same as the decay
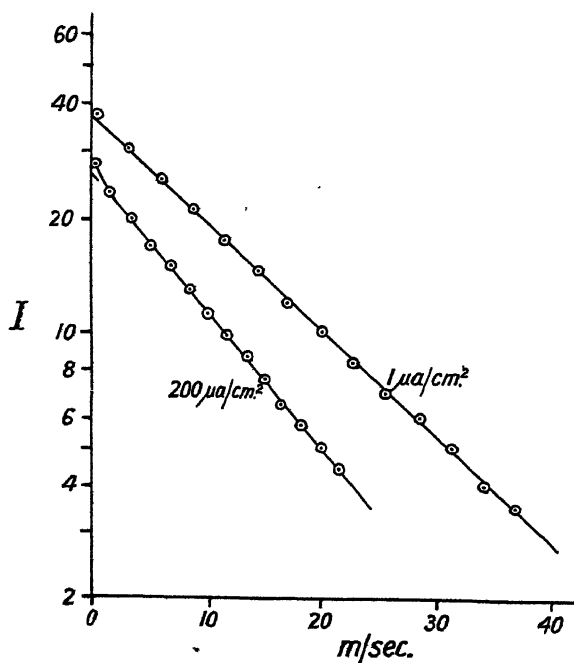
Figure 5. Decay process for Zn_2SiO_4.Mn : semi-logarithmic plots. Light units are arbitrary, and different for the two cases.

constant at very low current densities, but increased rapidly with current density, though scarcely at all in the linear manner to be expected (equation (4)). Figure 6 shows the changes for Zn_2SiO_4.Mn. The growth processes consisted of rapid and slow parts, though the constant given refers to the slow β part. The values for

Table 3. Decays in manganese class

Phosphor	$\mu a./cm^2$	β constant (sec.$^{-1}$)	% α process
$Zn_2B_2O_5$. Mn	1–300	63– 76	2–12
Zn_2SiO_4 . Mn	1–200	67– 78	0–23
ZnO . BeO . SiO_2 . Mn	1–400	98–123	8–25
$ZnAl_4O_7$. Mn	2–900	130–144	0– 4
ZnS . Mn	2–600	625–645	0–26

120 $\mu a./cm^2$ show, for example, that only half the light output comes from the β process. If α and β processes are assumed to be of equal efficiency in light production from electronic energy, then the current density producing the constant of 160 is not 120, but 60 $\mu a./cm^2$ This correction would considerably straighten the curve of the growth constant, making its rise nearly linear with

current density (see figure 6, *d*).   In contrast to this, the decay constant changed by less than 20% over the whole current-density range of the diagram, emphasizing the importance of current density in growth processes as required by theory.
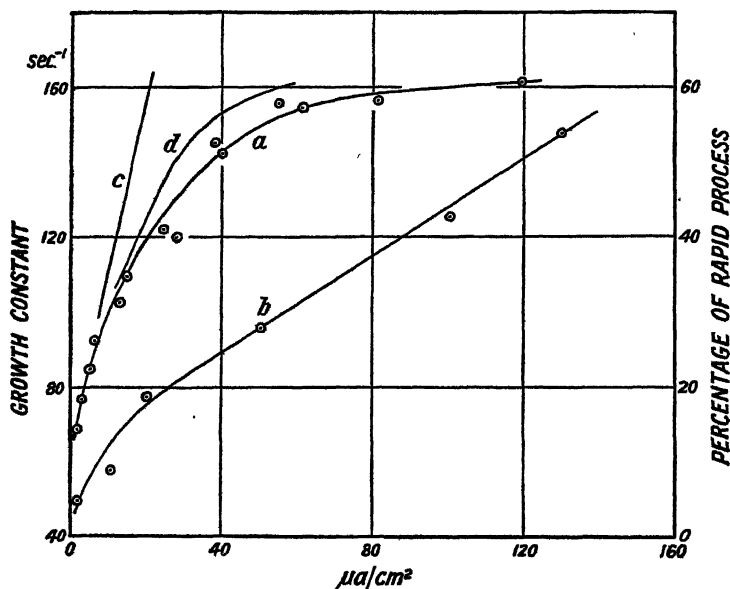
Figure 6.   Change of growth processes for Zn_2SiO_4. Mn with current density.

       a. Slow growth constant observed.
       b. Percentage of rapid process observed.
       c. Original slope of $(a) = A'i + B$ from equation (4).
       d. Corrected form of (a) : see text.

       If $B = 67$, $A' = 4\cdot3$ when i is measured in $\mu a./cm^2$.

In general the manganese class materials had a growth process well matched by a single exponential at low current densities. At increased current densities a good match could be made by two exponentials superimposed.

Effect of activator content

Interesting results were obtained by varying the amount of activator in zinc silicate.   Above about 1%, manganese produces absorption bands in the visible region and a marked change of decay rate, shown in table 4.

Table 4.   Change of decay constant of Zn_2SiO_4. Mn with Mn content

% Mn	Decay constant (sec.$^{-1}$)	Peak of emission (A.)	Relative efficiency
0·05	67	5245	8·4
0·2	67	5250	10·0
0·5	67	5255	9·7
1·0	80	5263	6·0
2·0	113	5291	5·5
5·0	550	5333	1·3

Very similar results for Zn_2SiO_4. Mn were reported by Fonda (1939) using ultra-violet radiation.

Effect of temperature

With similar arrangements to those described in the measurements on the Ag class, the effect of reducing the temperature of the phosphor to $-180°$ c. was examined for Zn_2SiO_4. Mn and ZnS . Mn. Only very slight differences in the decay constants were observed and they were too close to experimental error to be significant. Measurements on Zn_2SiO_4.Mn were also made at temperatures up to 250° c. : the growth constant seemed to be little affected by the change, but there was an increase of about 15% in the decay constant. As for the silver class, there is no such change as would be expected for processes involving ionization.

Survey of the more complex materials

An increase in the sensitivity of the apparatus made it possible to separate different parts of the emission spectrum in some cases and show that they had different decay constants.

In place of the photocell unit previously used, two new photoelectric multipliers were obtained, one with a maximum sensitivity in the red and the other in the

Table 5

Material	Emission spectrum	Current density (μacm^2)	Decay constants (sec.$^{-1}$)
Mg_2SiO_4 . Mn	Red band	2–500	17–23 ( )
			(α), 10–37%
	Blue band	2–500	c. 3000
Al_2O_3 . Cr	Red lines	50–400	210–250 (β)
	Green band	50–400	c. 2500
$ZnAl_4O_7$. Cr	Red lines	5–400	60 (β)
			c. 1000 (α)
	Green band	5–400	1400–3000
$ZnAl_4O_7$. Co	Red and blue bands	5–400	200–800
			1500–5000
$ZnAl_4O_7$. Sm	Orange lines	5–400	400 (β)
	Blue band	5–400	2000

blue-green. With these it was possible to examine materials hitherto too dull for test, and also, with the help of suitable filters, to separate different spectral regions.

The most interesting material was $ZnAl_4O_7$.Cr, which has an emission spectrum consisting of (1) a green band with a rapid exponential decay and a constant increasing from 1400 to 3000 sec.$^{-1}$ with increasing current density; and (2) a group of red lines with an exponential constant of 60 sec.$^{-1}$. With increase of current density the proportion of the green component increased, and at the same time a rapid red process appeared with a constant of about 1000 sec.$^{-1}$. At $-190°$ c. the

green also increased in relative proportion, the constant being apparently unchanged; the red was too weak to permit a determination. It appears that processes typical of both the Ag and Mn classes may occur in the same phosphor. It is not clear whether both types can be attributed to the one activator or whether the rapid decay regions, for example the higher frequency bands in $Zn\,Al_4O_7$. Cr and Mg_2SiO_4. Mn, are due to the base material. Zinc silicate, which when pure has a decay of the Ag type, and when activated by Mn joins the Mn class, might by careful control of the degree of activation yield a mixed phosphor of this type.

The different results are listed in table 5. They include a group of materials with $Zn\,Al_4O_7$ as a base and a number of different activators; this is of interest in expanding the Mn group of activators by Co, Cr, and Sm. The case of Sm needs further examination, for it will be seen that $CaO.Sm$ appears among the results of the Ag class in table 1.

In the last two materials, separation of spectral regions was incomplete owing to low luminous efficiencies.

(c) *Processes lasting more than 1 second*

So far no direct evidence for ionization processes has appeared in these results. The lack of marked temperature dependence suggests that ionization plays only a subsidiary part. On the other hand it seems certain that the primary process responsible for the absorption of the energy of the incident electrons must be one of ionization. There is direct evidence of its occurrence in some phosphors, particularly under photo-excitation: temperature-dependent decays of a hyperbolic form have been measured, and photoconductive effects have been observed in cases where no such decays were apparent, for example in ZnSe (Miller and Strange, 1938); and they are also reported for materials such as Zn_2SiO_4.Mn (Randall and Wilkins, 1939). To round off the work already described, it was thought important to measure ionization processes in phosphors under electronic bombardment even if present in very small proportions only. In the work described it was not possible to make the comprehensive survey of the subject that was intended. All that was achieved was the development of a new method of measurement suitable for examining phosphorescent decays when present in amounts less than 1% of the total initial intensity, and the application of this method to a few materials over a limited range of current density.

The general lay-out of the apparatus was very similar to that given in figure 3 except that (1) the cathode-ray monitor tube had the usual two pairs of plates instead of four pairs; (2) in place of the matching unit a timing unit for calibrating the rate of scan of the monitor tube was included; (3) the time base was extended to several seconds in a single sweep across the monitor tube, and (4) the H.T. unit of the multiplier had high output and low impedance in order to avoid the effects of changes in voltage during the initial sudden drop of light intensity.

The device was operated as follows :—With appropriate current in the C.R. tube (adjust P 5), a value of R_1 and a suitable filter for the multiplier were selected to give a convenient deflexion on the monitor within the previously determined linear range. I_0 was thus recorded in relation to a base line adjusted to the same position by the bias control P 2 after each change of sensitivity of the multiplier. To record

a decay curve starting from, say, 5% of I_0, the value of R_1 was increased by 20 times or the filter suitably changed. Depression of K caused the monitor to trace the decay curve on a y-scale 20 times as large as that for I_0, and still in the linear range. The decay curve was suitably disposed on the screen by adjustment of the scanning time, which was subsequently determined by calibration time pips recorded on a base line with the C.R. tube switched off. The y-axis was traced by alteration of P2. The timing unit was free-running, hence the time signals on the x-axis could not be accurately synchronized with the depression of K, but the error was very small. The traces were read off from a scale on the monitor tube face, or photographed and measured later.
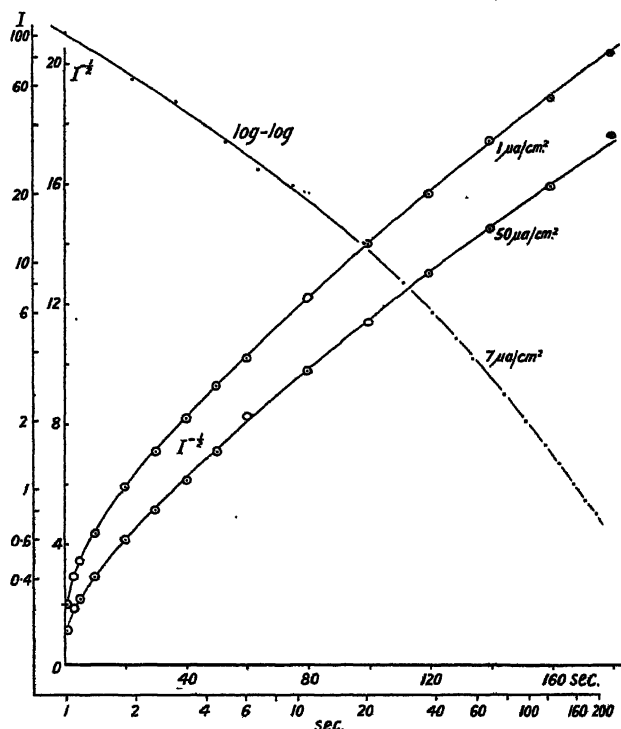
Figure 7. Decay of ZnS.Cu. Light units are arbitrary, but the same for the decays at 1 and 50 μa./cm.? Logarithmic scales refer to decay at 7 μa./cm.?

In most cases the traces for decay of ZnS.Cu and other materials of long afterglow gave a linear relation between time and $1/\sqrt{I}$ over the range $\frac{1}{2}$ to 4 seconds. It has since appeared that this period is too short to establish a bi-molecular decay, as the extended results of figure 7 show. These decay curves for ZnS.Cu were taken at 3 kv. on an elaborated version of the apparatus, in which a cyclic scan showed successive portions of the same decay curve one below the other, continuing for any desired length of time. From figure 7 it seems doubtful if any part of the decay process can be accurately classed as bimolecular, and even if we assume a power law in which $m \neq \frac{1}{2}$ (equation (1)), the logarithmic plot in figure 7 shows no finality in the slope up to 3 minutes' decay. A more thorough investigation of this part of the problem is evidently required.

§4. CONCLUSIONS

The experimental work shows that processes of exponential form occur widely in growth and decay of light output from inorganic phosphors. The random nature of the processes is further emphasized by the small effects of temperature change; that is, the experimental processes are not bimolecular changes which have become pseudo-exponential by excess of one "reactant", as suggested by Fonda (1939) for part of the willemite decay process at low temperatures.

The simple theory fails to account for the complexity of the results, especially in the changing proportions and speeds of the simultaneous processes which were measured. There was no definite evidence in favour of bimolecular decay, though non-exponential processes were found to be present in the growth at low current densities and in the decay at long times after excitation.

The phosphors could be divided into two classes, known as *silver* and *manganese* classes, distinguished primarily by the faster growth and decay processes of the former, which contained "pure" phosphors besides those activated by Ag, Cu, Bi, Zn and, possibly, Pb, Au. The manganese class included phosphors activated by Mn, Sm, Cr and Co. A summary of phosphor properties is given in the third paper of this series.

REFERENCES

BECQUEREL, 1861. *Ann. Chim. Phys.*, 62, 5.
FONDA, 1939. *J. Appl. Phys.*, 10, 408.
DE GROOT, 1939. *Physica*, 6, 275, 393.
LENARD *et al.*, 1928. *Handbuch d. Experimentalphysik, Leipzig*, 23, 176.
MILLER and STRANGE, 1938. *Proc. Phys. Soc.*, 50, 374.
NELSON *et al.*, 1939. *J. Appl. Phys.*, 10, 335.
PRINGSHEIM, 1939. *Trans. Faraday Soc.*, 35, 87.
RANDALL and WILKINS, 1939. *Nature, Lond.*, 143, 978.
RANDALL, WILKINS and GARLICK, 1945. *Proc. Roy. Soc.*, A, 184, 347–433.
SCHLEEDE and BARTELS, 1938. *Z. tech. Phys.*, 19, 364.
STRANGE, 1938. *Trans. Faraday Soc.*, 35, 95.
STRANGE, 1940. *Thesis* (London University).

APPENDIX

Description of figure 3

The apparatus includes—

(1) A key K which, when pressed, raises the potential of point A from 76 v. to 170 v., at the same time disconnecting the valve anode from the 360 v. supply.

(2) A scanning unit B consisting of a condenser charged through a MSP4 valve, acting as a constant-current device. Its operation is controlled by the potential of point A which, through potentiometer P3, determines the potential of the grid of the valve. The rate of scan is controlled by P3.

(3) A cathode-ray tube C, which may be demountable, containing the screen under test. The cathode of the electron gun is connected to point A and the modulator to potentiometer P5, which, with the key raised, determines the current through the tube. Depression of the key makes the cathode 94 v. positive and cuts off the current.

(4) A photoelectric multiplier D with an output stage consisting of two KTZ41 valves in push-pull, giving a linear response of 100 v. either positive or negative in sense.

(5) A matching unit E consisting of condensers and resistances in series connected to the 360-v. supply and the key. With the key up, the voltage across the condensers is the difference between A and 360. Depression of the key causes the two sides of the condensers to be "shorted" through their respective resistances; the potential falls exponentially in each case.

(6) A special monitor tube F with four sets of electrostatic deflexion plates used for observing the results of mixing signals from the photoelectric unit and the matching unit.

CATHODO-LUMINESCENCE: PART II. CURRENT SATURATION AND VOLTAGE EFFECTS

By J. W. STRANGE and S. T. HENDERSON,

Electric and Musical Industries, Ltd., Hayes, Middlesex

ABSTRACT. The light output from phosphors under steady electron beams has been measured at constant voltage and varying current density, and thus "current saturation" has been found to vary greatly in extent for different materials. Similarly the change of light output with varying voltage at constant current density has shown different characteristics for different phosphors, but without the expected variations on changing the current density.

§1. INTRODUCTION

THE preceding paper on cathodo-luminescence (referred to as Part I) was concerned with the variation of light output with time from phosphors under a steady electron beam, or decaying after excitation. We now consider the variations of light output at *equilibrium excitation* when the current density or the voltage of the impinging electrons is altered.

In the early days of experimental work on cathodo-luminescence, Lenard (1903) proposed for the light output from a phosphor the law

$$L = ki(V - V_0)^q, \qquad \qquad \dots\dots(7)$$

where k is a constant, i is the current, V is the voltage of the impinging electrons, V_0 is a "dead voltage" typical of the phosphor and attributed (Lenard and Saeland, 1909) to damaged surface layers, and $q = 1$. This law has been found to have limited application only. The fact that in general L increases less rapidly than i has long been known (as *current saturation*), while several workers have attempted to fit their results to expressions like equation (7) in which q has some value, not necessarily integral, between 1 and 3. A square law has frequently been used in this way (Brown, 1937; Nottingham, 1937, 1939).

These and other investigators have studied another property of phosphors, namely secondary emission characteristic, which by its variability adds to the difficulty of accurate light-output measurements. In general these materials

show a ratio of less than one secondary electron to each primary one up to an incident energy of, say, 100 v. For greater energies, up to several thousand volts, the ratio is greater than one; then at a point commonly called the second cross-over the ratio falls to one, and for still higher beam voltages the phosphor tends to charge down towards cathode potential and maintain the voltage difference of the second cross-over. Surface contamination caused, for example, by tube baking, or by burning of the phosphor under excessive bombardment, are two causes of changes in the second cross-over point. Other experimental difficulties met in the accurate measurement of the light output at current densities above a few μa./cm? are:

(1) the raising of the temperature of the phosphor, with consequent change in efficiency;

(2) the burning of the phosphor, resulting in the destruction of its fluorescent properties, particularly under steady excitation at high current densities;

(3) the non-uniform distribution of current in the exciting electron beam produced by the usual electron-gun systems.

Most of the published results for the light output from fluorescent screens have been made using a scanning electron beam, i.e. moving at speeds of about 10^5 cm./sec. across the screen. It is very difficult to define accurately the conditions used in such an operation and to interpret the results obtained.

Results of measurements under steady excitation have been published by Nottingham (1937), who used an evacuated, sealed-off tube containing six phosphors deposited on glass plates fixed close to a central electrode, which could be used for measuring the current. The improved method described in his later paper (1939) still appeared to be open to doubt on the points raised above, especially regarding surface contamination which affected the second cross-over point.

§2. RELATION OF LIGHT OUTPUT TO CURRENT DENSITY AT CONSTANT VOLTAGE

The apparatus used consisted of the demountable tube shown in figure 8. With a rapid pumping system this worked satisfactorily at low current densities, though at high current densities the difficulties listed above were encountered, as in the case of sealed-off tubes. However, by the use of a new ballistic method these difficulties were successfully avoided.

A thermionic cathode of the impact type, (Ehrenberg, 1939), in which a tantalum cap is heated to about 2000° c. by electronic bombardment, yielded the necessary high currents. The beam, rendered roughly parallel by the magnetic coil N, was defined by the electrodes A, B and C, of which A prevented secondary electrons moving from B to C, B defined the beam falling on the sample and was kept about 9 v. negative to A and C, while C collected secondary electrons from the sample D for beam-current measurement. A and C were at the final anode potential, B had to be in good contact with the tube wall, leaving no gaps, but separated from the graphite coating E as shown.

The glass plate carrying the eight sample phosphor screens was mounted on a ground joint in the glass end-plate. The samples could thus be brought into the same position under the electron beam in turn. The end-plate was sealed on with Apiezon compound Q.

Preliminary tests were made to confirm the accuracy of beam-current measurement by means of secondary emission to electrode C; also to check the current distribution across the electron beam, which varied by less than 10% at low current densities, but became less uniform at increased currents, with the effect of spreading out changes in slope of the (light *vs*. current density) curve.

The beam-current measurement was made with a mirror galvanometer F with universal shunt, giving sufficient sensitivity for steady or ballistic measurements· The light measurement employed a similar galvanometer G with a shunt, receiving the output from a photo-cell H and bridge-type D.C. amplifier K. This ampli-
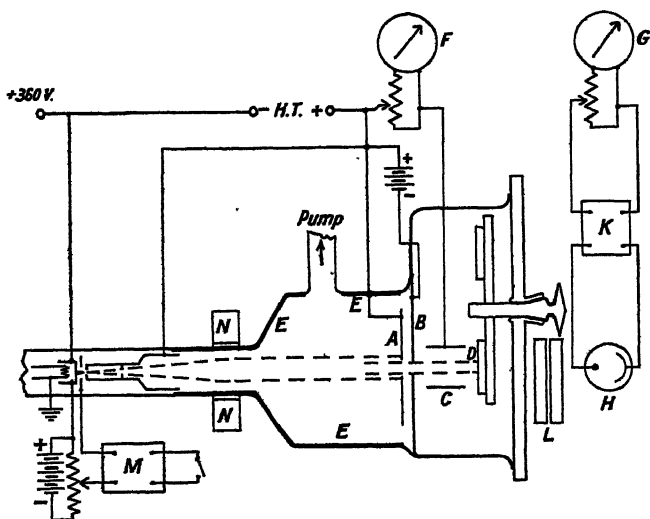
Figure 8. Apparatus for light-output measurements.
Electron paths shown by broken lines.

fier used two LP2 triodes operating over the lower part of their characteristics, and gave a very steady and linear output when properly balanced. Colour or neutral filters L were used as required.

Ballistic measurements were made by applying a potential pulse to the cathode modulator system of the electron gun for a predetermined time, by means of the multivibrator M, and measuring simultaneously the photo-cell output and the beam current, each galvanometer being critically damped instead of freely oscillating. Control experiments showed this method to be reliable (cf. Masius, 1925). The pulse length in ballistic measurements was fixed at 0·1 second, which reduced the error due to finite growth period of the light output to about 2% for willemite and less for other phosphors. The upper limit of the pulse was determined by the possibility of overheating or burning the samples. The absence of serious effects was checked by using a screen of $CaWO_4$, of which the light output is very sensitive to temperature changes.

Experimental results

Measurements made on a wide variety of materials showed that in all cases there was a linear relation between light output and current at very low current densities, as indicated by equation (5) (Part I). This fact was employed in all subsequent measurements to eliminate various factors irrelevant to the main purpose of the tests. Among these may be mentioned the variation in the intrinsic brightness and in the position of the emission band, differences in thickness of the screen and in the light absorption in the materials. With the original linear slope as a basis for comparison, and taking it as unity in all cases, such variables could be eliminated. The efficiencies given will thus be relative to the efficiency of the initial linear range. It may be noted that the absolute efficiency reported for the most efficient materials is of the order of 10% (Leverenz and Seitz, 1939).

Three materials were selected for detailed examination: $ZnS.Ag$ which is typical of the group of rapidly decaying phosphors called the *Ag class* in Part I, and $ZnS.Mn$ and $Zn_2SiO_4.Mn$, which both belong to the *Mn class* of phosphors. These last two have been found at low current densities to have growth and decay processes approximating to those expected for a phosphor in which the luminescence is a simple random process. The light output should follow equation (4), Part I, if the theory is correct. These materials were also chosen because in $ZnS.Mn$ and $ZnS.Ag$ we have the same base materials with different activators and in $ZnS.Mn$ and $Zn_2SiO_4.Mn$ the same activator with different base materials. The comparison should throw further light on the importance of these differences.

The three materials were tested at an incident voltage of 5000 over a current range up to $300 \mu a./cm^2$? The relative efficiencies determined by the method described above are given in table 6; these figures do not imply that all the materials were equally efficient at the lowest current densities.

Table 6.   Variation of light output with current density
(Relative efficiencies)

Current density ($\mu a./cm^2$)	1	50	100	150	200	250	300
$ZnS.Ag$	100	100	100	100	100	97	93
$ZnS.Mn$	100	86	72	62·5	57	53·5	50
$Zn_2SiO_4.Mn$	100	68	50	40·5	36	32	30
,, calculated	94	25	14	10·5	8	6	5

The calculated values for $Zn_2SiO_4.Mn$ given in table 6 were derived from the simple theoretical expression for light output and the values of A' and B previously found (Part I, equation (5) and figure 6). The failure of these calculated values to agree even approximately with experiment was to be expected from the disagreement already shown in figure 6, where the growth constant failed to conform to the linear variation with current density which the theory predicted, and this failure persisted even after correction for the proportion of rapid process (see Part III).

The efficiency figures in table 6 seem to indicate fundamental differences in properties between the phosphors. The influence of methods of preparation

was investigated, and here only minor changes were observed. Zn_2SiO_4. Mn made by four different methods gave values of 31, 34, 41 and 44 for the relative efficiencies at 250 $\mu a./cm^2$. A similar group of ZnS. Ag materials gave 90, 93, 95 and 97.

On the other hand, marked changes were observed when the activator content was changed in Zn_2SiO_4 beyond the range of about 0·2 to 1·0% already remarked upon in the decay measurements (Part I). Using the same set of Zn_2SiO_4. Mn materials with manganese varying from 0·05 to 5·0%, light-output curves were obtained in which, over the range 0·2 to 2·0% Mn, there was a progressive increase in the length of the initial linear part. Above and below these proportions the change in form was rapid, with the slope observed at 250 $\mu a./cm^2$ lowest for 0·05% Mn, about equal for 0·2 to 2% Mn, and still greater for 5% Mn.

Measurements on the three principal materials were made up to current densities of 1 ma./cm², but considerable difficulty was experienced in ensuring reproducible conditions. The results can be summarized thus for ZnS. Mn and Zn_2SiO_4. Mn: the rate of change of slope observed over the current range 20–200 $\mu a./cm^2$ lessened at higher current densities, and the curves tended again towards a linear relation. There appeared to be no indication of an approach to a maximum value of the light output as suggested by the simple theory.

§3. VOLTAGE EFFECTS

(a) *Growth and decay processes*

It seemed possible that the probability constant for excitation, A (or A'), introduced in Part I, might vary with the density of exciting radiation, and thus with the voltage of the impinging electrons. Using Zn_2SiO_4. Mn, experiments were made in the apparatus of §3 (b), Part I, to measure the growth constant, which the simple theory requires to be $Ai+B$. Some additional difficulty was experienced in obtaining satisfactory current distribution across the electron beam falling on the sample, owing to the wide range of current density and voltage used. Thus the results will not show the maximum changes that may in fact occur. Table 7 A shows the usual increase of growth constant with current density, but no appreciable effect by change of voltage; table 7 B shows that the initial rapid growth process decreases in amount with voltage increase at any current density (except the lowest used). Table 8 gives comparable results for the proportion of rapid decay process, α. Other materials behaved similarly, for example $ZnAl_4O_7$. Mn and $Zn_2B_2O_5$. Mn.

Some connexion may be seen between some of the changes in growth and decay process so far recorded. In Part I it appeared that increased current density at constant voltage caused increases in proportion of rapid growth and decay processes. In the present case we have varying voltage, and, since faster electrons penetrate more deeply, the average energy dissipation per second per unit volume of phosphor is greater for smaller incident voltage at constant power input. If we consider both cases in terms of power input it is seen that in each an increased amount of rapid process accompanies increased energy dissipation per second per unit volume. Compare for example the groups of figures marked in tables 7 B and 8: the percentages are greater for lower voltage at equal power input. Current

Table 7.   Growth process for Zn_2SiO_4. Mn at different voltages and current densities

A. Constants for main process (sec.$^{-1}$)

Current density ($\mu a./cm^2$)	Voltage (kv.)						
	1	2	3	4	5	6	8
1	72	78	79	79	81	73	70
10	110	111	114	118	118	108	111
20	121	132	130	122	125	122	122
50	137	147	139	132	137	143	143
100	139	139	139	135	134	133	137

B. Percentage of rapid process

Current density ($\mu a./cm^2$)	Voltage (kv.)						
	1	2	3	4	5	6	8
1	9	15	11	9	6	15	22
10	25	17	14	9	3	5	4
20	33	20	17	26	17*	9	10
50	44	33*	35	38	25	24†	25
100	56*	46	47†	46	38	38	38

Note: Four independent groups of figures have been selected: those marked * in table 7 B (0.1 watt/cm^2), those marked † in table 7 B (0.3 watt/cm^2), and similar groups in table 8 (* 0.2, and † 0.8 watt/cm^2).

Table 8. Percentage of rapid decay process (α) for Zn_2SiO_4. Mn at different voltages and current densities

Current density ($\mu a./cm^2$)	Voltage (kv.)						
	1	2	3	4	6	8	10
0·3	10	8	10	5	5	5	2
2·5	8	10	6	6	2	0	0
20	20	12	10	8	6	8	10*
50	25	20	17	12*	12	15	15
100	25	16*	18	17	14	18†	18
200	—	—	27	26†	22	22	20

saturation may be similarly treated in terms of energy, and it is shown in Part III of this series that a greater percentage loss of efficiency occurs for the lower voltage if the power input is kept constant.

(b) *Value of voltage exponent q*

If, as is generally assumed, the rate of energy dissipation in depth varies with the speed of the electrons at the particular depth considered, it might be expected that under different current-saturation conditions different values of q (equation (7))

would be found. A further series of experiments was directed towards the determination of q at various current densities. The apparatus was similar to that described in §2. Certain modifications were called for by the need to maintain similar conditions of current distribution at widely varied voltages, whereas previously the voltage was kept constant (except in the work described in §3 (a)). Unless this uniformity is maintained, the variations in light output due to current density changes may easily obscure the effect of voltage changes. It was also desirable to increase the current density as much as possible above the previous maximum of 300 μa./cm.^{2}, to emphasize the expected changes of q with current density.

The electron gun consisted of a cathode of the impact type (Ehrenberg, 1939), a modulator, an accelerator, and an anode containing an aperture of 0·1 inch diameter. The modulator was used only for applying the controlling pulse and was at cathode potential during the period of excitation. The current through the tube was then dependent on the temperature and emissivity of the cathode and the potential of the accelerator, with only slight dependence on the anode voltage. No focusing of the electron beam was attempted, as this would obviously alter with anode voltage; in consequence the size of the beam striking the screen was also independent of the voltage and was determined only by the size of the aperture in the anode cylinder and its distance from cathode and screen.

The usual method of operation was to vary the current in the tube by altering the cathode temperature. The latter was difficult to control with precision, and it was necessary to take a number of readings scattered about the required current density value and construct from them a small section of the current density *vs.* light-output curve at that voltage. The value of the light output at the required current could then be read off with a high degree of accuracy. The same ballistic method, with pulses of 0·1 second, was used as in the experiments of §2.

Experimental results

Some of the results for the three main materials, Zn_2SiO_4. Mn, ZnS. Mn and ZnS. Ag are given in figures 9 to 11. The most striking of these are the results for the first material. Over the wide range of voltage from 300 to 9000 v., and without any adjustment for V_0, there was no variation in the value of q, and over a considerable change of current density, from 5 to 550 μa./cm.^{2}, the variation was very small.

The other two materials needed corrections for V_0 to straighten the tail of the curve. With the extension of the measurements to voltages close to these values of V_0, the corrections could be made with certainty. The low values of V_0 are evidence that the screens suffered a minimum of contamination. ZnS. Mn showed some instability at high currents and voltages, and a low second cross-over of the secondary emission curve, but otherwise both sulphides behaved like Zn_2SiO_4. Mn over a wide voltage range, with a similar small, if not negligible, change of q with current density.

Besides the failure of these materials to show any marked change of q with current density, there appeared quite definitely the fact that q is characteristic of the phosphor and not the same for all materials. The possibility that this variation was due to the absorption of light in the phosphor was considered. It was,
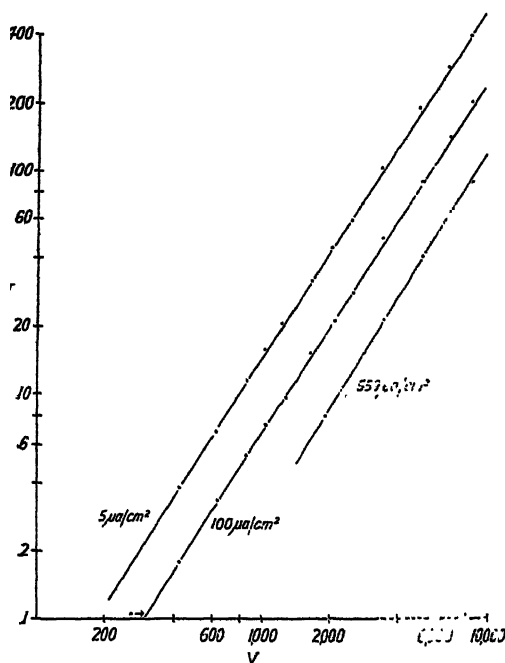
Figure 9. Voltage-brightness relation for
Zn$_2$SiO$_4$.Mn. Light units different
for different current densities.
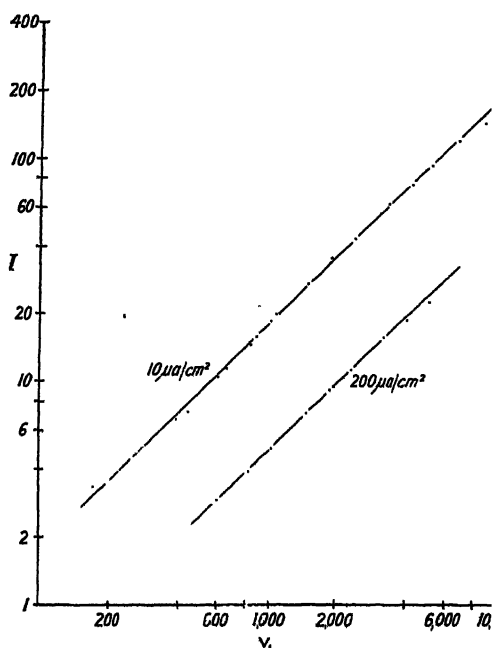
Figure 10. Voltage-brightness relation for
ZnS.Mn. Light units different for
different current densities.
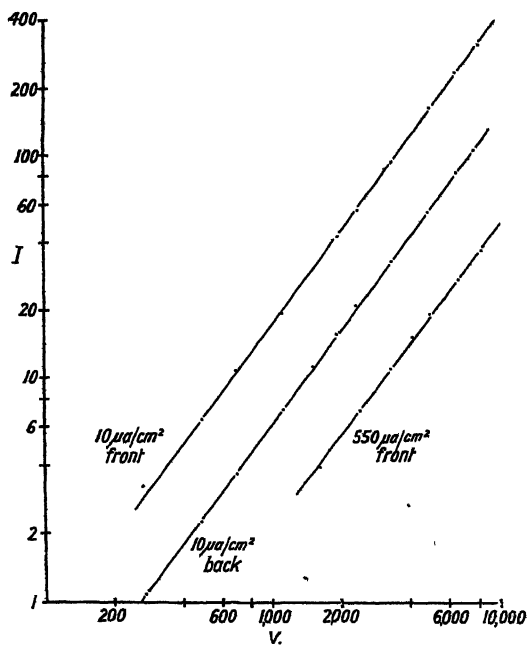
Figure 11. Voltage-brightness relation for ZnS.Ag.
Light units different for different current densities.

however, ruled out by measuring the light emitted from the back (i.e. the side facing the electron source), and from the front of the screen for ZnS . Ag. There was no difference between the slopes derived from the two sets of figures. Some such effect would be expected, perhaps only in a much higher voltage range.

The results are collected in table 9.

. Table 9. Values of q and V_0

Current density ($\mu a./cm^2$)		5	10 front	10 back	100	200	550
Zn_2SiO_4 . Mn	q V_0	1·53 0			1·54 0		1·59 0
ZnS . Mn	q V_0		1·0 200 v.			0·98 250 v.	
ZnS . Ag	q V_0		1·36 150 v.	1·36 150 v.		.	1·32 150 v.

The theoretical aspects of the work recorded in this paper, and in the previous one on growth and decay processes, are considered in the third paper of this series.

ACKNOWLEDGMENTS (PARTS I AND II)

REFERENCES

BROWN, 1937. *J. Opt. Soc. Amer.*, **27**, 186.
EHRENBERG, 1939. British Patent 523,240.
LENARD, 1903. *Ann. Phys , Lpz.*, **12**, 449.
LENARD and SAELAND, 1909. *Ibid.*, **28**, 485.
LEVERENZ and SEITZ, 1939. *J. Appl. Phys.*, **10**, 479.
MASIUS, 1925. *Phys. Rev.*, **25**, 211.
NOTTINGHAM, 1937. *J. Appl. Phys.*, **8**, 762 ; 1939. *Ibid.*, **10**, 73.

CATHODO-LUMINESCENCE:
PART III. DISCUSSION OF RESULTS

BY J. W. STRANGE AND S. T. HENDERSON,
Electric and Musical Industries, Ltd., Hayes, Middlesex

ABSTRACT. The interpretation of the experimental results in the preceding two papers is attempted, but cannot be adequate without more knowledge about electron absorption in phosphors. Some possibilities are outlined with regard to the mechanism of luminescence. There is some evidence in favour of a new type of voltage absorption law.

§ 1. INTRODUCTION

THE two earlier papers in this series have given the experimental facts of the present study, and this paper will examine theoretical matters, commencing with qualitative aspects of luminescence and proceeding to a more quantitative discussion of some of the data. We may first enquire what bearing the results have on current views of the mechanism of luminescence, considering mainly the evidence from the growth and decay processes. This evidence has already been shown to be too complicated to fit the simple theory of the first paper (referred to here as Part I), and there is also no clear agreement between the theory and the observations on light emission at equilibrium excitation. A more thorough knowledge of the processes of electron absorption in crystals is required before much progress can be made in interpreting the results: this need will appear again later when empirical quantitative methods are introduced to deal with the voltage results. Recent developments in the theory of trapping states in phosphors may provide a valuable new approach to some of the unsolved problems. So far, however, these ideas of Randall and his co-workers (1945) seem to be applicable mostly to decays of a much longer order, produced by photo-excitation at energy densities of much lower value, than those used in the present work.

§ 2. QUALITATIVE ASPECTS OF LUMINESCENCE

The distinctions between excitation by ultra-violet radiation and by an electron beam are well known. Photo-luminescence depends on the presence of absorption bands which may be of different kinds in the same phosphor, giving different decay rates (Kröger, 1939), and possibly different luminous efficiencies. In cathodo-luminescence all atoms of the phosphor within the range of incident electrons must take part in the primary absorption of energy. This may cause excitation without ionization of (a) the atoms of the base material, or (b) the impurity "activator" atoms; or it may produce ionization of both kinds of atom. Ionization is generally agreed to be the more probable result, and where the activator content is small, as in the *Ag class*, far more atoms of the base material than of the activator must undergo this process. Since luminous efficiencies are of the order of 10% in cathodo-luminescence, and are similar for materials of widely

differing activator contents ($0 \cdot 01 \%$ in ZnS . Ag, 1% in Zn_2SiO_4 . Mn), and since the active centre determines the frequency of the emission band or lines, it is a necessary conclusion that energy absorbed from the electron beam by the base material is transferred to the activator atoms before luminescence occurs. This transfer may be effected in two ways: either by movement of electrons in the conduction band, and at the same time of positive holes in levels below the ground state in the active centres, or, on the other hand, by an " exciton " mechanism (Mott and Gurney, 1940, p. 208). The first alternative would be supported by the presence of photo-conductivity, which is found in some, but not all, phosphors. The second is in agreement with the small amount of very slow phosphorescence observed in most cases, for it implies that the final excitation of the centre will not be an ionization, and therefore exponential forms may be expected for growth and decay processes.

The lack of temperature dependence in growth and decay (Part I) is decidedly in favour of the hypothesis that the light-emission processes do not directly involve ionized electrons, but this cannot be taken as a conclusive proof. From the converse point of view, well defined non-ionizing processes may sometimes show a type of temperature dependence resembling the " freezing in " of luminescence in materials where ionization precedes emission. For example, in $ZnAl_4O_7$. Cr the transition of a $3d$ electron in Cr^{+++} gives a doublet ^4F—^2G near 6870 A. which has satellite lines due to intramolecular Stark effect. The lines of greater frequency than the doublet are much reduced in relative intensity by fall of temperature, and almost disappear at $-180°$ c. The lines of smaller frequency than the doublet are not greatly affected in intensity, relative to the main lines, by temperature change between $-180°$ c. and $300°$ c. It would be of interest, though difficult, to determine growth and decay constants and luminous efficiency for separate lines in this emission spectrum. Again, pseudo-exponential processes may be temperature dependent (Part I, § 4).

The evidence of growth and decay measurements suggests the presence of more than one activator level (or group of levels), with different characteristic speeds, their relative participation in the process of luminescence varying with current density. Shift of emission-band position with altered density of excitation radiation would support this view, though conversely the absence of spectral shift does not rule out the possibility of more than one level. Certain small effects have been found on changing from electron excitation to the less intense irradiation by ultra-violet (Henderson, 1939), but an increase of current density by 40 times in an electron beam at 5 kv. has been found to produce no change in the peak position of the ZnS . Ag band, nor has the increase of voltage from 2 to 10 kv. at constant current density significantly altered the position of the ZnS . Zn band.

With regard to emission bands, it is probable that some small amount of energy transfer occurs by optical absorption. When a phosphor has two emission bands, one may be of sufficiently high frequency to excite the other in a separate screen, or in layers of the same screen further from the exciting radiation, for example in ZnS . Cu or ZnS . CdS . CuAg. The absorption bands concerned are not usually intense enough to make the effect important.

The assumption of two simultaneous emission processes of exponential form reduces to some extent the disparity between the simple theory of Part I and the

experimental results. The calculated brightness figures in table 6, Part II, are for the slow β process: the difference from the experimental values may then be taken to represent the contribution of the rapid α process, if both α and β are equally efficient in light production. Hence by calculation the percentage of rapid process increases from 6% at 1 μa./cm^2 to 63% at 50 μa./cm^2 and 78% at 200 μa./cm^2. These percentages are increased if the non-linear values of the growth constant are taken from figure 6, Part I. The observed proportions of rapid growth and decay processes were, however, far less than these calculated values. Also, inconsistent values of A' were found by calculation from the experimental data for light output, growth, and decay respectively. This modification of the simple theory is, therefore, of no great use in reconciling theory with experiment, at any rate for Zn$_2$SiO$_4$.Mn.

The effect of collisions of the second kind has been considered as a possible mechanism to account for increasing growth and decay rates with increasing current density. If the collision of an electron and an excited centre were to cause an immediate transition with luminescence, a kind of acceleration effect would appear as the electron density increased.. The most noticeable result would be a nearly instantaneous fall of light output at the end of the excitation period owing to the cessation of this acceleration. The fall would increase with current density.

There is in fact a very rapid fall at high current densities under scanning conditions. Screens of Zn$_2$SiO$_4$.Mn and ZnS.Mn were scanned by a spot of high current density ($\eqsim$5 amp./cm^2), when some proportion of the decay occurred in a process lasting not more than a few microseconds. This may be regarded merely as the increased amount of rapid decay (α process, see table 2, Part I); alternatively it could be attributed to an acceleration effect, but for the fact that there is a similar rapid rise at the start of excitation by high current densities, which would not be expected as a consequence of the effect in question. The experiments of Part I showed the growth and decay curves to be always similar in form if their ordinates (light output) were measured from the initial value, zero or I_0 respectively. There was no suggestion of the dissimilarity which would arise from an acceleration process. A more plausible explanation of the very fast decay at high intensities has been given by Mott and Gurney (1940, p. 217), who believe that the lifetime of the excited states determines the decay rate under these conditions.

A more detailed mechanism is required to explain the changes of speed with current density in the separate α and β processes. If it is assumed that the lifetime of excited metastable states in the activator atoms may be of the same order as the observed decay times (cf. Mott and Gurney, p. 209), the following scheme seems to be worth consideration.

The main excitation occurs, presumably, by ionization of atoms of the base material, followed by the recombination of ions with electrons and the generation of exciton waves. The excitons raise electrons of the activator atoms from the ground state X to non-ionized metastable levels Y. X and Y are located, according to the current theory of luminescence, in a forbidden energy zone between the highest full level F of the matrix atoms and the normally empty conduction band. Luminescence arises from the transition Y→X, which gives exponential forms largely independent of temperature changes. The separation of X and Y must,

in effect, have a Gaussian error distribution to account for emission band width and shape, though this variation must not affect the transition probabilities.

Assume now that the X-Y group of levels may take up varying positions in the forbidden region, which may happen for activator atoms differently situated in the lattice. It seems likely that the probabilities of excitation X→Y and emission Y→X will vary according to these positions. That is, the energy difference F − X will affect the probabilities. Constancy of spectral emission will be assured if the separation X − Y remains the same for any value of F − X, apart from the differences causing band width. There are thus available a number of transitions of different effective speeds. For the experimental conditions of excitation over relatively long periods of time, this should result in increased rates of growth and decay with increased density of excitation, owing to saturation of the more slowly decaying states.

This preferential saturation should not be observed for times of excitation short compared with the decay times, unless another condition is imposed. Experiments quoted above showed that in scanning, with excitation times of about a microsecond for each screen element, increases occurred in decay rates similar to those under saturation excitation. The extra condition necessary to account for this is that increased probability of excitation must be accompanied by decreased probability of emission between the levels X and Y. The slower transitions Y→X will then be favoured at low excitation densities for any length of pulse. It is not evident whether the changes in probability would occur, in the sense mentioned, by increase or by decrease of the energy separation F − X.

The mechanism described could provide the variations with current density of a single process, e.g. the β process in ZnS . Ag. It would appear to need duplication at a higher range of excitation and emission probabilities to explain the simultaneous α process, and it is not certain if metastable transitions within the activator atoms of the *Mn class* can be entirely responsible for the slow exponential β process observed in cases where the decay constant is of the order of 10^2 sec.$^{-1}$.

Non-exponential processes, mainly in long-term decay, must have their rates determined by the behaviour of ionized electrons and the presence of traps, but this type of luminescence, though it has practical applications, is of minor importance in electronic excitation.

§3. DECAY RATES AND CURRENT SATURATION

This aspect of the experimental results can be considered without reference to the mechanism of electron absorption (see §4). Increase of energy absorption per second per unit volume of phosphor leads to increased overall rates of growth and decay. This appears most simply in the results for constant incident voltage (Part I). Similarly, power increase at constant incident voltage causes loss of light efficiency, but there is no clear relation between this change and the effects on growth and decay. The two kinds of effect may be independent and not occur necessarily over the same range of power input. Thus in the case of ZnS . Ag, a linear (light output *vs.* current density) curve between 0 and 200 μa./cm.^{2} at 5 kv. is accompanied by great increases in speed of growth and decay, while Zn$_2$SiO$_4$. Mn exhibits much smaller changes in speed but a considerable fall in efficiency. Alternatively, the increased speed in ZnS . Ag may be the reason for the unaltered

efficiency, simply by making the centres more rapidly available for re-excitation. This idea is supported qualitatively by considering phosphors in the order ZnS.Ag, ZnS.Mn, Zn_2SiO_4.Mn. This is the order of increasing degree of current saturation, and also of decreasing speeds. The experimental values of A' at $i \simeq 0$ were 33, 15, and 4·3 sec.$^{-1}\mu a.^{-1}$ respectively; of B, approximately 10^4 to 10^5, 625, and 67 sec.$^{-1}$ respectively. These figures suggest that the comparatively large value of B is one reason for the lack of saturation effects in ZnS.Ag in the current range of the experiments.

§4. QUANTITATIVE ASPECTS

The processes of electron absorption and light emission in phosphors depend on several factors which might be measurable, but on which we have insufficient information at present:

(a) the manner in which electron velocities (or voltages) decrease with the thickness of matter traversed;

(b) the effective electron current, which may vary through the slab;

(c) the relation between absorbed energy and light output for any volume element of the phosphor, which involves the method of transmission of energy through the crystal as well as the efficiency of conversion to light at the active centres.

Apart from these major factors there are:

(d) scattering and straggling in the electron beam, which must be of some importance but will not be discussed; and

(e) light absorption and scattering, which occur in phosphor layers, but seem to have a negligible effect on the results (see Part II, table 9).

(a) Current saturation

In Part II, §3 (a), some correlation was obtained by use of the energy parameter instead of separate V and i values. The current-saturation data for Zn_2SiO_4.Mn may be examined in a similar way. Owing to the method of plotting current density *vs.* light output on relative efficiency scales, no values were available for direct comparison of the curves for different voltages. When plotted with the same initial slope, the 4 kv. and 10 kv. curves were very similar in shape over the range 0 to 600 $\mu a./cm^2$. The 4kv. curve is plotted in figure 12, a. The 10 kv. curve can be calculated from this, using the results of table 9, Part II, since brightness at 10 kv. = (brightness at 4 kv.) × $(10/4)^q$. The values of q for Zn_2SiO_4.Mn in table 9 give a linear relation between current density and $(10/4)^q$: from this relation the appropriate values at each current density were taken to derive curve b, figure 12. If the experimental curve for 10 kv. is fitted to this theoretical curve, starting from point A, the other marked points are obtained, which show a satisfactory agreement to exist between the data of two different types of experiment. Next, a and b are replotted on abscissae representing incident energy per second instead of current density, which gives curves c (4 kv.) and d (10 kv.).

The increased saturation at lower voltage is now evident. There is a nearly

constant ratio of 2·67 between the values of energy input rate at any relative efficiency in the range considered. Thus the power input required to cause a given loss of luminous efficiency is nearly proportional to the voltage used.

So far no assumption has been made about the distribution of the energy through the slab of phosphor. If the electron current is considered to remain constant during the absorption process, then it must be concluded that the voltage change in absorption is of the type mentioned in Part II, §3 (b), by which a given depth of matter produces a greater relative loss of speed in slow electrons than in fast ones. The Thomson-Whiddington law is the classical formulation of this process for fast electron absorption in gases and metal foils. It may be given as

$$y = a(V_1^2 - V^2), \qquad \qquad \ldots \ldots (8)$$
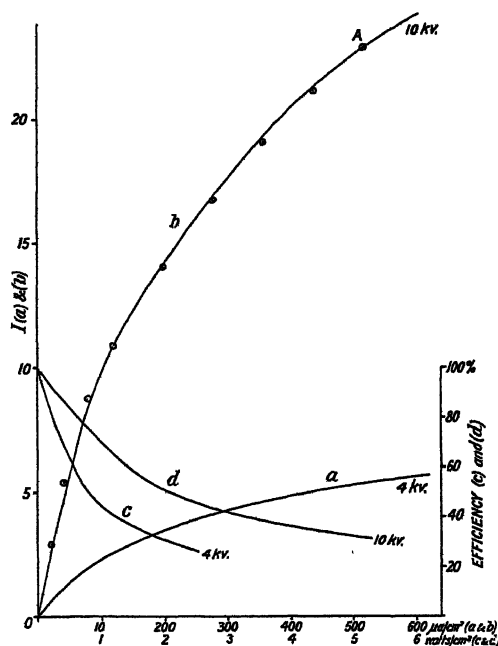
Figure 12. Current saturation in $Zn_2SiO_4 . Mn$.
Light units are arbitrary (curves *a* and *b*). Efficiency scale refers to curves *c* and *d*.

where *y* is the depth of penetration at which electrons of original voltage V_1 have been retarded to (a most probable) voltage V, and *a* is a constant for the absorbing material. In the comparatively slow speed range used in the present work, the value of *a* may decrease with V_1 (Allen, 1932), but apart from this correction it is not certain that equation (8) is applicable at all to the case of absorption in phosphor crystals. The discussion of decay processes in terms of power input, in Part II, §3 (a), is valid for any voltage absorption law since the deeper penetration of faster electrons distributes the incident energy over a greater volume of phosphor, and the decay rates change appropriately. In the next section it will appear that the Thomson-Whiddington law, or a similar one with indices greater than unity, leads to difficulties in interpretation.

(b) *Voltage exponent q*

The light output agrees experimentally with the expression $k(V_1 - V_0)^q$, where k, V_0, and q are constants. Explanations are required for:

(1) the constancy of q for a given material over a wide range of current density, showing no significant dependance on current saturation effects;

(2) values of q which differ from unity;

(3) differing values of q for similar materials.

These facts are more suitable for quantitative treatment than the current-saturation data. As before, the assumptions made will be rather speculative, in the absence of precise knowledge on electron absorption.

Current saturation and q

Provided that current saturation phenomena are ignored, and if the current density i is taken to be constant through the phosphor layer, simple energy considerations show that any voltage absorption law will make q equal to unity. Various other cases have been worked out for the Thomson-Whiddington law, with the additional assumption of current variation, and in each case q was unity whether i decreased linearly with thickness traversed or by power laws analogous to equation (8).

If current saturation is taken into account, the theoretical effect is to increase q markedly at higher excitation densities, contrary to what was found by experiment. To overcome this difficulty various assumptions may be made. For instance, suppose that energy dissipation is uniform through the phosphor, and that the power expended per unit volume depends only on the initial current i_0. For these conditions, the Thomson-Whiddington law would give $q=2$, while a linear decrease of voltage with penetration, or

$$y = b(V_1 - V) \qquad \qquad \ldots\ldots(9)$$

would give $q=1$, and current saturation would not affect these values. In general, to obtain $q=n$, the total energy dissipated must vary as V_1^n, hence the case of $q=1$ is the only probable one from simple energy considerations. This cannot be adduced as strong evidence for equation (9) because of the lack of support for the hypothesis of uniform energy dissipation.

A different assumption regarding energy dissipation, less rigid than the one just discussed, leads to the same independence of current saturation, and incidentally supports equation (9) as against equation (8). Consider a layer of phosphor of thickness y_1 in which electrons of the original voltage are just absorbed. Let dE/dy be the rate of dissipation of energy at a depth y in this layer. Now

$$\frac{dE}{dy} = i \cdot \frac{dV}{dy} + V \cdot \frac{di}{dy},$$

and from the particular modes of variation of i and V chosen, a curve may be drawn for the relation

$$\frac{dE}{dy} \text{ vs.} y.$$

Its area between $y=0$ and $y=y_1$ gives the total energy absorbed. Variations of

dE/dy with y mean that current-saturation effects may occur, but these need not alter the value of q if the overall degree of current saturation is the same for all incident voltages. This requires that for any incident voltage the same fraction of the total energy shall be dissipated at any given value of dE/dy.

A simple example is the linear decrease of both V and i with y, both V and i becoming zero together. Then dE/dy decreases linearly to zero with distance penetrated in the layer, and $q=1$. Other variations of i with y or V give the same result, provided the linear variation of V remains, as in equation (9).

If, on the other hand, the Thomson-Whiddington law is accepted, the elimination of differential saturation effects is not obtained with any of the proposed simple variations of i.

This brief discussion shows that the observed constancy of q under current-saturation conditions can be accounted for, provided that equation (9) holds, while no such simple explanation has been found with a power law like equation (8). This appears to throw some doubt on the validity of the Thomson-Whiddington law, at any rate for the voltage range of the experiments. Some independent support for the linear type of voltage variation is afforded by Fano's calculations on electron scattering (1940), which require the average penetration of slow electrons to be nearly proportional to their energy.

Voltage exponent q not unity

So far the only reasonable theoretical value of q not affected by current saturation appears to be unity; in practice, values of q greater than unity were obtained under conditions of current saturation.

A non-florescent surface layer on the phosphor, whether of foreign matter or of decomposed phosphor, would produce an apparent increase in the value of q by reducing the incident voltage more in proportion for slow electrons than for fast ones. This effect could not be responsible for the observed values of q, since the "dead voltage" corrections (V_0 in Part II, table 9) were much too small to account for changes in q of the order of 0·5.

Hitherto it has been assumed that light output is a function of dE/dy alone. It is doubtful if this position can be maintained in face of the results in question. It is likely that the apparent abnormalities arise from subsequent stages of the luminescence, following the primary absorption of energy, and that this cannot be treated by the methods outlined above.

From this point of view Fano (1940) has attempted a theoretical explanation of the work of Brown (1937), who found $q=2$ for $Zn_2SiO_4.Mn$ between 200 and 1300 v. at low current densities. Fano examined electron scattering and diffusion, the resulting average depth at which excitons (or "wandering excitations") were produced for any voltage, and the probability of non-luminescent transfers of exciton energy at crystal surfaces. This is a much more complex treatment than the one we have given, and it is doubtful if it can be expected to give accurate q values, especially as many assumptions and approximations are made. The value of Fano's work lies not so much in its support of Brown's evidence, which is open to some criticism on the grounds of probable screen contamination, but in showing a method of approach, and indicating the possibility of values of q other than unity.

Different q values for similar phosphors

Events in the luminescent process subsequent to the primary absorption must also be responsible for the differences observed between the values of q for ZnS . Ag and ZnS . Mn.   There is a difference in crystal composition between these phosphors, apart from the activators, the content of blende being probably from 20 to 40% in the ZnS . Ag, with the remainder wurtzite, while the ZnS . Mn contained 95% or more wurtzite.   Differences of exciton diffusion and absorption by the active centres, as well as differences in surface properties, might well occur here.   It would be of interest to measure q for a range of ZnS . Ag samples in which differing heat treatment had developed varying proportions of the two crystal forms.

To summarize this section on the voltage exponent: the absorption law $y = b(V_1 - V)$ will account for the observed lack of dependence of q on current saturation, but no simple explanation can be made on the basis of equation (8). The processes by which the energy of the primary electrons is conveyed to the luminescent centres must be held responsible for the fact that q is not unity in some cases, while differences of crystal structure are the probable cause of the characteristic values observed.   Current variations through the phosphor, which may have a profound effect on the results, have so far not been determined, and the experimental difficulties of such measurements would obviously be very great.

The experimental conditions may not have given the simplest results, and it would perhaps be better to use electron beams of much higher velocity, and very thin screens which only slightly retard the beams, together with means of avoiding secondary emission trouble.

§5. SUMMARY AND CONCLUSION

The similarities existing among all the phosphors tested are more remarkable than their differences.   Those with Ag, Cu, Zn, Bi (? Pb, Au) as activators, and the "pure" phosphors, exhibit two exponential processes in growth and decay of luminescence, with comparatively high constants.   The relative proportions of the processes change rapidly with current density, while the values of the "constants" change to a smaller degree.   The activators all have atoms with completed electronic shells.   Their concentration in phosphors should be small, of the order of 0·01%.   The emission bands in this *Ag class* are wide and structureless and little affected by temperature; in general the wider the band, the faster the growth and decay processes, and the more rapid the decrease of light output efficiency with temperature increase.

Phosphors in the *Mn class* have as activators Mn, Cr, Co, with incomplete inner electronic shells.   The exponential growth and decay processes vary slowly in speed and relative proportions with current density, and the constants are much lower than in the *Ag class*.   Activator contents are high for Mn (0·1 – 1·0%), lower for Cr and Co.   Emission bands are narrow, or consist of broadened lines; some resolution into lines is often observed at low temperatures.   Temperature dependence of efficiency is generally much less than in the *Ag class*, but current-saturation effects are more prominent for the conditions used.

Samarium occupies an intermediate position as an activator.   Its constants are high for the *Mn class* (table 1, Part I), though its other properties fit best into this group.

The exponential processes require a more complex explanation than is provided by the usual mechanism assumed for crystal luminescence. It is suggested that the idea of electron traps might be applied to these phenomena. Energy dissipation has proved to be a useful parameter in correlation of results in different sections of the work, but for its adequate exploitation requires a knowledge of voltage and current changes in phosphors which we do not possess.

The simple hypothesis of linear voltage change in absorption accounts for some of the experimental results: in view of the complicated nature of luminescence, even so far as it is understood, such apparently simple explanations are not to be accepted too readily. There is a need for more experimental work on the lines of that recorded in these papers, and a more fundamental study of the electron-absorption process in phosphors.

REFERENCES

ALLEN, 1932. *J. Franklin Inst.*, **214**, 25.
BROWN, 1937. *J. Opt. Soc. Amer.*, **27**, 186.
FANO, 1940. *Phys. Rev.*, **58**, 544.
HENDERSON, 1939. *Proc. Roy. Soc.*, A, **173**, 323.
KRÖGER, 1939. *Physica*, **6**, 369, 764, 779.
MOTT and GURNEY, 1940. *Electronic Processes in Ionic Crystals* (Oxford Univ. Press).
RANDALL, WILKINS, GARLICK, 1945. *Proc. Roy. Soc.*, A, **184**, 347–433.

THE INTEGRAL BREADTHS OF DEBYE-SCHERRER LINES PRODUCED BY DIVERGENT X RAYS

By A. J. C. WILSON,
University College, Cardiff

ABSTRACT. Divergence of the incident x-ray beam produces appreciable phase differences between different parts of a crystal, even in the size range for which line broadening occurs. The broadening due to this phase difference is calculated. It is ordinarily negligible, as the increase in the width of the line is of the order of $1 \cdot 6 \, t \cos \theta$, where t^3 is the volume of a crystal and θ is the Bragg angle. The special case of film and source equidistant from the crystal is investigated in greater detail.

§1. INTRODUCTION

IN the theoretical treatment of the diffraction of x rays by crystals it is usual to consider a parallel incident beam. In practice, however, the rays diverge from a source a few centimetres from the specimen. The source may be real, the focus of the x-ray tube, or effective, part of the slit system of the camera. X-ray wave-lengths are sufficiently small for the divergence to introduce appreciable phase differences between different parts of a crystal small enough to produce

line broadening. It seemed desirable, therefore, to investigate the diffraction of divergent x rays by crystals of this order of size, in case the divergence should have an appreciable effect on the line broadening.

It is found that the broadening on the film is of the order of $1 \cdot 6 \, t \cos \theta$, where t^3 is the volume of a crystal and θ is the Bragg angle. To this approximation it is independent of the radius of the camera and the wave-length of the x rays. The broadening is thus negligible for crystals small enough to have their sizes measured by line broadening ($\sim 10^{-5}$ cm.), and it is not practically significant for the larger crystals used to give comparison lines. Debye–Scherrer lines would still have a finite breadth, due to the non-homogeneity of the incident x rays and imperfections in the experimental arrangement, even if diffraction broadening were entirely absent. Divergence broadening of the comparison lines would be indistinguishable from broadening due to the finite diameter of the specimen, and would disappear in the elimination of the broadening due to experimental conditions by methods such as that of Jones (1938).

The mathematical treatment is greatly simplified in the special case in which the source and the film are equidistant from the specimen. The calculation of the integral breadths of reflections from spherical crystals is carried through in detail for this case, so that the transition from small-particle (diffraction) broadening to large-particle (divergence) broadening can be followed. For any particular angle of reflection the integral breadth, β, is a minimum for a particle size given by $t = 1 \cdot 00 (Q\lambda)^{\frac{1}{2}} / \cos \theta$, where Q is the camera radius and λ is the wave-length of the x rays. The broadening on the film, $Q\beta$, is $1 \cdot 40 (Q\lambda)^{\frac{1}{2}}$. This is too small to be detected with normal technique.

§ 2. DIFFRACTION OF DIVERGENT X RAYS

In figure 1, the source of the x rays is at O, P is a vector joining the source to the centre of gravity of the crystal, Q is a vector from P to the point on the circumference of the camera at which it is desired to calculate the reflected intensity, r_j is a vector from P to the jth unit cell of the crystal, and 2θ is the angle of deviation.
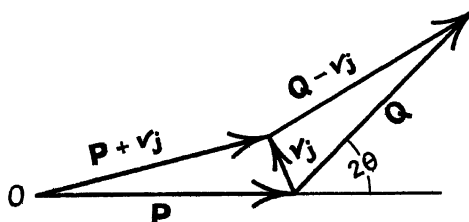
Figure 1.

The path difference between the rays diffracted from P and P + r_j is

$$|P + r_j| - |P| + |Q - r_j| - |Q|$$

$$= \{(P + r_j) \cdot (P + r_j)\}^{\frac{1}{2}} - P + \{(Q - r_j) \cdot (Q - r_j)\}^{\frac{1}{2}} - Q$$

$$= P\left\{\left[1 + \frac{2P \cdot r_j}{P^2} + \frac{r_j \cdot r_j}{P^2}\right]^{\frac{1}{2}} - 1\right\} + Q\left\{\left[1 - \frac{2Q \cdot r_j}{Q^2} + \frac{r_j \cdot r_j}{Q^2}\right]^{\frac{1}{2}} - 1\right\}$$

$$= P\left\{\frac{\mathbf{P}\cdot\mathbf{r}_j}{P^2} + \frac{\mathbf{r}_j\cdot\mathbf{r}_j}{2P^2} - \frac{\mathbf{P}\cdot\mathbf{r}_j^2}{2P^4} + \ldots\right\} + Q\left\{-\frac{\mathbf{Q}\cdot\mathbf{r}_j}{Q^2} + \frac{\mathbf{r}_j\cdot\mathbf{r}_j}{2Q^2} - \frac{\mathbf{Q}\cdot\mathbf{r}_j^2}{2Q^4} + \ldots\right\}$$

$$= \mathbf{r}_j\cdot(\mathbf{p}-\mathbf{q}) + \left(\frac{1}{2P} + \frac{1}{2Q}\right)\mathbf{r}_j\cdot\mathbf{r}_j - \mathbf{r}_j\cdot\left(\frac{\mathbf{pp}}{2P} + \frac{\mathbf{qq}}{2Q}\right)\cdot\mathbf{r}_j, \qquad \ldots\ldots(1)$$

where $\mathbf{p}$ and $\mathbf{q}$ are unit vectors in the direction of $\mathbf{P}$ and $\mathbf{Q}$. Let $\mathbf{p}-\mathbf{q}=\lambda(\mathbf{h}+\mathbf{\eta})$, where $\mathbf{h}$ is the vector of the reciprocal lattice which makes $|\mathbf{\eta}|$ smallest. Then $\mathbf{r}_j\cdot\mathbf{h}$ is an integer and $\mathbf{r}_j\cdot\mathbf{h}\lambda$ can be dropped from the expression for the path difference. Let

$$\sigma = \frac{1}{2P\lambda} + \frac{1}{2Q\lambda}, \quad \delta = \frac{1}{2P\lambda} - \frac{1}{2Q\lambda}, \quad \rho=\sigma/\delta, \quad \text{and} \quad \mathbf{\Psi} = \frac{\mathbf{pp}}{2P\lambda} + \frac{\mathbf{qq}}{2Q\lambda}, \qquad \ldots\ldots(2)$$

where $\mathbf{\Psi}$ is a dyadic. The amplitude of the reflected beam is then (the "structure factor" being omitted for brevity)

$$G = \Sigma_j \exp\{2\pi i(\mathbf{r}_j\cdot\mathbf{\eta} + \sigma\mathbf{r}_j\cdot\mathbf{r}_j - \mathbf{r}_j\cdot\mathbf{\Psi}\cdot\mathbf{r}_j)\} \qquad \ldots\ldots(3)$$

$$\doteq V^{-1}\int \exp\{2\pi i(\mathbf{r}\cdot\mathbf{\eta} + \sigma\mathbf{r}\cdot\mathbf{r} - \mathbf{r}\cdot\mathbf{\Psi}\cdot\mathbf{r})\}\,dv, \qquad \ldots\ldots(4)$$

where V is the volume of one unit cell and the integration is over the volume of the crystal. This integral may be simplified somewhat by a proper choice of axes. Let

$$\mathbf{a}=(\mathbf{p}-\mathbf{q})/2\sin\theta, \quad \mathbf{b}=(\mathbf{p}+\mathbf{q})/2\cos\theta, \quad \mathbf{\gamma}=\mathbf{p}\times\mathbf{q}/\sin2\theta. \qquad \ldots\ldots(5)$$

It will readily be seen that $\mathbf{a}$, $\mathbf{b}$, $\mathbf{\gamma}$ are unit and orthogonal. Then

$$\mathbf{\Psi} = \sigma\sin^2\theta\,\mathbf{aa} + \delta\sin\theta\cos\theta\,\mathbf{ab}$$
$$+ \delta\sin\theta\cos\theta\,\mathbf{ba} + \sigma\cos^2\theta\,\mathbf{bb}. \qquad \ldots\ldots(6)$$

It is desired to find unit and orthogonal axes $\mathbf{\alpha}$, $\mathbf{\beta}$ such that

$$\mathbf{\Psi}\cdot\mathbf{\alpha}=A\mathbf{\alpha}, \quad \mathbf{\Psi}\cdot\mathbf{\beta}=B\mathbf{\beta}, \qquad \ldots\ldots(7)$$

where A and B are constants. It can be verified that

$$\left.\begin{array}{l}\mathbf{\alpha}=\dfrac{\mathbf{a}+\{|\rho\cot2\theta| - |(\rho^2\cot^2 2\theta+1)^{\frac{1}{2}}|\}\mathbf{b}}{\sqrt{2}\,|\{1-|\rho\cot2\theta(\rho^2\cot^2 2\theta+1)^{\frac{1}{2}}| + \rho^2\cot^2 2\theta\}^{\frac{1}{2}}|}, \\[4mm] \mathbf{\beta}=\dfrac{\mathbf{a}+\{|\rho\cot2\theta| + |(\rho^2\cot^2 2\theta+1)^{\frac{1}{2}}|\}\mathbf{b}}{\sqrt{2}\,|\{1+|\rho\cot2\theta(\rho^2\cot^2 2\theta+1)^{\frac{1}{2}}| + \rho^2\cot^2 2\theta\}^{\frac{1}{2}}|},\end{array}\right\} \qquad \ldots\ldots(8)$$

$$\left.\begin{array}{l}A=\tfrac{1}{2}\{\sigma - |(\sigma^2\cos^2 2\theta + \delta^2\sin^2 2\theta)^{\frac{1}{2}}|\}, \\[2mm] B=\tfrac{1}{2}\{\sigma + |(\sigma^2\cos^2 2\theta + \delta^2\sin^2 2\theta)^{\frac{1}{2}}|\},\end{array}\right\} \qquad \ldots\ldots(9)$$

and that $\mathbf{\alpha}\to\mathbf{a}$ and $\mathbf{\beta}\to\mathbf{b}$ as $\delta\to0$. With axes in the directions $\mathbf{\alpha}$, $\mathbf{\beta}$, $\mathbf{\gamma}$,

$$\mathbf{r}=x\mathbf{\alpha}+y\mathbf{\beta}+z\mathbf{\gamma}, \quad \mathbf{\eta}=\xi\mathbf{\alpha}+\eta\mathbf{\beta}+\zeta\mathbf{\gamma},$$

$$G=V^{-1}\int \exp\{2\pi i[x\xi+y\eta+z\zeta+\sigma(x^2+y^2+z^2)-Ax^2-By^2]\}\,dv$$

$$=V^{-1}\int \exp\{2\pi i(x\xi+y\eta+z\zeta+Bx^2+Ay^2+\sigma z^2)\}dv \qquad \ldots\ldots(10)$$

$$=V^{-1}\int \exp\{2\pi i[(B^{\frac{1}{2}}x+\xi/2B^{\frac{1}{2}})^2-\xi^2/4B]\}dx \int \exp\{2\pi i[(A^{\frac{1}{2}}y+\eta/2A^{\frac{1}{2}})^2$$
$$-\eta^2/4A]\}dy \int \exp\{2\pi i[(\sigma^{\frac{1}{2}}z+\zeta/2\sigma^{\frac{1}{2}})^2-\zeta^2/4\sigma]\}dz. \qquad \ldots\ldots(11)$$

This integral is difficult to evaluate for an arbitrary crystal shape, but for a sphere whose radius a is large compared with $B^{-\frac{1}{2}}$, $A^{-\frac{1}{2}}$, $\sigma^{-\frac{1}{2}}$, it may be evaluated approximately by the use of properties of Fresnel integrals. It is well known that

$$\left|\int_{x_1}^{x_2}\exp\{2\pi i c x^2\}dx\right|^2 \begin{cases} \sim(2c)^{-1} & \text{for } x_1 \text{ and } x_2 \text{ of opposite sign and} \\ & \qquad x_1{}^2,\ x_2{}^2 \gg c^{-1}, \\ \sim 0 & \text{for } x_1 \text{ and } x_2 \text{ of the same sign.} \end{cases} \quad \dots\dots(12)$$

For a sphere, the limits of integration of the three integrals in equation (11) are $\pm(a^2-y^2-z^2)^{\frac{1}{2}}$, $\pm(a^2-z^2)^{\frac{1}{2}}$, $\pm a$ respectively. The modulus of the value of the first integral in equation (11) is, therefore, approximately $(2B)^{-\frac{1}{2}}$ for $\xi^2<4B^2\{a^2-y^2-z^2\}$ (i.e. for $y^2<\{a^2-z^2-\xi^2/4B^2\}$), 0 otherwise. Similarly the modulus of the product of the first two integrals is $(4AB)^{-\frac{1}{2}}$ for $\eta^2<4A^2\{a^2-z^2-\xi^2/4B^2\}$ (i.e. for $z^2<\{a^2-\xi^2/4B^2-\eta^2/4A^2\}$) and zero otherwise, and the modulus of the triple integral is $(8AB\sigma)^{-\frac{1}{2}}$ for $\zeta^2<4\sigma^2\{a^2-\xi^2/4B^2-\eta^2/4A^2\}$ (i.e. for $\xi^2/4B^2+\eta^2/4A^2+\zeta^2/4\sigma^2<a^2$) and zero otherwise. Then the intensity of reflection as a function of η is

$$H=VGG^* = \frac{1}{8AB\sigma V},\quad \frac{\xi^2}{4B^2}+\frac{\eta^2}{4A^2}+\frac{\zeta^2}{4\sigma^2}<a^2,\quad \dots\dots(13)$$
$$=0,\ \text{otherwise.}$$

In other words, the intensity of reflection is approximately constant and equal to $(8AB\sigma V)^{-1}$ within an ellipsoid of semi-axes of lengths $2Ba$, $2Aa$, $2\sigma a$ and directions α, β, γ, approximately constant and equal to zero outside. The total intensity I is equal to $(8AB\sigma V)^{-1}$ times the volume of the ellipsoid:

$$I= \tfrac{4}{3}\pi.2Ba.2Aa.2\sigma a.(8AB\sigma V)^{-1}= \tfrac{4}{3}\pi a^3 V^{-1},$$

which is the total number of unit cells, as it should be. The intensity with θ between θ and $\theta+d\theta$ is proportional to the volume of the ellipsoid contained between two planes perpendicular to a and a distance $2\cos\theta d\theta/\lambda$ apart. The maximum intensity is, therefore (figure 2),

$$\left(\frac{dI}{d\theta}\right)_0 = \frac{2\cos\theta}{\lambda}.H.\pi.2\sigma a.T$$
$$= \frac{\cos\theta}{2\lambda}.\frac{\pi a T}{ABV},\quad \dots\dots(14)$$

where T is the semi-diameter of the ellipsoid in the direction of b. The integral breadth is therefore

$$\beta = \frac{2I}{(dI/d\theta)_0} = \frac{16\lambda a^2 AB}{3T\cos\theta}.\quad \dots\dots(15)$$

The explicit expression for T is rather complex; it may be shown to be

$$T=a\left\{\frac{\dfrac{\boldsymbol{\beta}\cdot\mathbf{a}^2}{\dfrac{\boldsymbol{\alpha}\cdot\mathbf{a}^2}{4B^2}+\dfrac{\boldsymbol{\beta}\cdot\mathbf{a}^2}{4A^2}}+\dfrac{\boldsymbol{\alpha}\cdot\mathbf{a}^2}{\dfrac{\boldsymbol{\beta}\cdot\mathbf{a}^2}{4B^2}+\dfrac{\boldsymbol{\alpha}\cdot\mathbf{a}^2}{4A^2}}}{}\right\}^{\frac{1}{2}}.\quad \dots\dots(16)$$

This simplifies considerably for $P=Q$. For most cameras this is approximately true, and the simplified expression will give some estimate of the integral breadth due to the divergence of the x-ray beam. The simplified expression is

$$T=2Aa,\quad \dots\dots(17)$$

and the integral breadth becomes

$$\beta = \frac{8\lambda aB}{3\cos\theta} = \frac{8a}{3}\cdot\frac{\cos\theta}{Q} = 1\cdot65t\frac{\cos\theta}{Q}. \qquad \ldots\ldots(18)$$

The actual broadening on the film is equal to the camera radius times the integral breadth, i.e. to $1\cdot65\,t\cos\theta$. This becomes appreciable when a is a few hundredths of a millimetre.
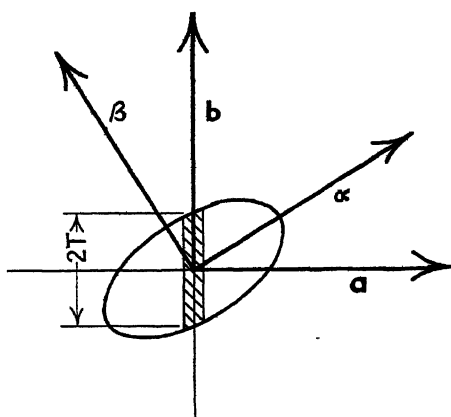

Figure 2.

The actual value of the numerical factor in equation (18) will depend on the shape of the crystal, and for non-spherical crystals on the indices of the reflection. Equation (11) can be evaluated similarly for other simple shapes; reflection from a face of a cubic crystal leads to a factor 2 instead of $1\cdot65$. Equation (18) should, however, give the order of magnitude of the effect.

§ 3. DETAILED CALCULATION FOR EQUIDISTANT SOURCE AND FILM

The whole calculation simplifies considerably for $P=Q$, so it is perhaps worth while to examine in more detail the transition from small-particle (diffraction) broadening to large-particle (divergence) broadening for a spherical particle in this special case. Equation (10) becomes

$$G = V^{-1}\!\int\exp\{2\pi i[x\xi+y\eta+z\zeta+\sigma(x^2\cos^2\theta+y^2\sin^2\theta+z^2)]\}dv, \quad \ldots\ldots(19)$$

where both x and ξ are measured in the direction of $\mathbf{h}$. The intensity of reflection as a function of η is

$$H = VGG^* = V^{-1}\!\int_v\!\int_{v'}\exp\{2\pi i\,L\}\,dv\,dv' \quad \text{if}$$

$$L\equiv[\xi(x-x')+\eta(y-y')+\zeta(z-z')+(\{x^2-x'^2\}\cos^2\theta$$
$$+\{y^2-y'^2\}\sin^2\theta+z^2-z'^2)] \qquad \ldots\ldots(20)$$

so that

$$\frac{dI}{d\theta} = \frac{2\cos\theta}{\lambda}\int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty}H\,d\eta\,d\zeta$$

$$= \frac{2\cos\theta}{\lambda V}\int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty}\!\int_v\!\int_{v'}\exp\{2\pi i[L]\}dv\,dv'\,d\eta\,d\zeta. \qquad \ldots\ldots(21)$$

The integrals with respect to η and ζ are singular, being zero if $y' \neq y$ or $z' \neq z$, and infinite for $y' = y$ and $z' = z$, but "by an appropriate limiting process" (Patterson, 1939 a, p. 973) it may be shown that the double integrals with respect to η and y', ζ and z' each have the value unity. Equation (21) becomes

$$\frac{dI}{d\theta} = \frac{2\cos\theta}{\lambda V} \iiiint \exp\{2\pi i[\xi(x-x') + \sigma(x^2 - x'^2)\cos^2\theta]\}\, dx\, dy\, dz\, dx', \ldots\ldots (22)$$

$$\left(\frac{dI}{d\theta}\right)_0 = \frac{2\cos\theta}{\lambda V} \iiiint \exp\{2\pi i\sigma\cos^2\theta(x^2 - x'^2)\}\, dx\, dy\, dz\, dx'. \ldots\ldots (23)$$

This may be expressed in terms of Fresnel integrals and cylindrical coordinates (axis in the direction of **h**) as

$$\left(\frac{dI}{d\theta}\right)_0 = \frac{2\cos\theta}{\lambda V} \int_0^{2\pi}\int_0^a \left[\int_{-(a^2-r^2)^{\frac{1}{2}}}^{(a^2-r^2)^{\frac{1}{2}}} \exp\{2\pi i\sigma\cos^2\theta x^2\}dx\right]$$

$$\left[\int_{-(a^2-r^2)^{\frac{1}{2}}}^{(a^2-r^2)^{\frac{1}{2}}} \exp\{-2\pi i\sigma\cos^2\theta x'^2\}dx'\right] r\, dr\, d\phi$$

$$= \frac{4\pi}{\lambda V\sigma\cos\theta} \int_0^a [C^2(2\sigma^{\frac{1}{2}}\cos\theta\sqrt{a^2-r^2}) + S^2(2\sigma^{\frac{1}{2}}\cos\theta\sqrt{a^2-r^2})]r\, dr$$

$$= \frac{\pi}{\lambda V\sigma^2\cos^3\theta} \int_0^{2\sigma^{\frac{1}{2}}a\cos\theta} [C^2(u) + S^2(u)]u\, du, \ldots\ldots (24)$$

where

$$C(u) + iS(u) = \int_0^u \exp\{\pi i u^2/2\}du.$$

The integral breadth is, therefore,

$$\beta = \frac{2I}{(dI/d\theta)_0} = 2 \cdot \frac{4\pi a^3}{3V} \cdot \frac{\lambda V\sigma^2\cos^3\theta}{\pi} \cdot \left[\int_0^{2\sigma^{\frac{1}{2}}a\cos\theta} [C^2(u) + S^2(u)]u\, du\right]^{-1}$$

$$= \frac{2\lambda}{3a\cos\theta} \cdot \frac{1}{D(2\sigma^{\frac{1}{2}}a\cos\theta)}, \ldots\ldots (25)$$

where

$$D(u) = 4u^{-4}\int_0^u [C^2(u) + S^2(u)]u\, du. \ldots\ldots (26)$$

To progress further it is necessary to evaluate the function $D(u)$. Since $C^2(u) + S^2(u) \to u^2$ as $u \to 0$, its value for small values of u is 1. For large values of u, $C^2(u) + S^2(u) \to \frac{1}{2}$, so that the asymptotic value of $D(u)$ is u^{-2}. A series convenient for small values of u may be obtained as follows. It is known (Preston, 1895, p. 276) that

$$C^2(u) + S^2(u) = M^2 + N^2, \ldots\ldots (27)$$

where M and N are known series satisfying the differential equations

$$\frac{dM}{du} = 1 - \pi u N, \qquad \frac{dN}{du} = \pi u M \ldots\ldots (28)$$

and, therefore,

$$\frac{d(M^2 + N^2)}{du} = 2M, \ldots\ldots (29)$$

and

$$D(u) = 2u^{-4}\int_0^{u^2}\left\{\int_0^{u^2} Mu^{-1}d(u^2)\right\}d(u^2). \ldots\ldots (30)$$

Integrating the series for M (Preston, p. 275) term by term gives

$$D(u) = 1 - \frac{2\pi^2 u^4}{3.4.1.3.5} + \frac{2\pi^4 u^8}{5.6.1.3.5.7.9} - \frac{2\pi^6 u^{12}}{7.8.1.3.5.7.9.11.13} + \cdots$$
$$\cdots \cdots (31)$$

The ratio of successive terms in this series is

$$- \frac{\pi^2 u^4 (2n-1)(2n)}{(2n+1)(2n+2)(4n-1)(4n+1)},$$

which approaches 0 for sufficiently large n for any value of u. It is therefore absolutely convergent for all values of u. It is, however, inconvenient for u greater than 2. Table 1 gives some numerical values of $D(u)$. Those for $u < 2$

Table 1. Values of $D(u)$

	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0	1·0000	0·99999	0·99982	0·99911	0·99720	0·99317	0·98590	0·97405	0·95624	0·93101
1	0·8969	0·8534	0·8000	0·7373	0·6670	0·5917	0·5151	0·4413	0·3749	0·3190
2	0·2753	0·2440	0·2226	0·2077	0·1951	0·1821	0·1669	0·1508	0·1371	0·1257
3	0·1181	0·1118	0·1062	0·0992	0·0921	0·0856	0·0811	0·0774	0·0738	0·0695
4	0·0653	0·0628	0·0595	0·0569	0·0540	0·0512	0·0491	0·0472	0·0452	0·0430
5	0·0413									

were calculated from the series, those for $u > 2$ by numerical integration of four-place tables of $C(u)$ and $S(u)$. In the range 0·5–2·0 the greatest difference between the values calculated by the two methods is 0·0003; the mean difference is about 0·0001.

With these values of $D(u)$, the integral breadth for any particular case can be calculated by equation (25). There is a minimum value of β which occurs for $D + uD' = 0$, i.e. for $u \doteqdot 1·24$. The corresponding value of a is

$$1·24/2\sigma^{\frac{1}{2}} \cos \theta \sim 40,000 \, \text{A. for } Q = 10 \text{ cm.}, \quad \lambda = 2 \text{ A.}, \quad \theta = 45°.$$

The minimum value of β is $1·40\sqrt{\lambda/Q}$; the actual broadening on the film, $Q\beta$, is $1·40\sqrt{Q\lambda} = 0·006$ mm., an amount undetectable with normal technique.

Equation (25) and the discussion in the previous paragraph have been given in terms of the radius of the particle, a. In terms of the cube root of its volume, t, the expressions become

$$\beta = \frac{1·0747\lambda}{t \cos \theta} \cdot \frac{1}{D(1·24\sigma t^{\frac{1}{2}} \cos \theta)}, \qquad \cdots \cdots (32)$$
$$t = 1·00/\sigma^{\frac{1}{2}} \cos \theta \text{ for minimum } \beta.$$

The minimum values of β and $Q\beta$ are unchanged. For $t/\sqrt{Q\lambda}$ small, equation (32) becomes $\beta = 1·0747\lambda/t \cos \theta$, in agreement with the results of Patterson (1939 b) and Stokes and Wilson (1942).

REFERENCES

JONES, F. W., 1938. *Proc. Roy. Soc.*, A, 166, 16.
PATTERSON, A. L., 1939 a. *Phys. Rev.*, 56, 972 ; 1939 b. *Ibid.*, 56, 978.
PRESTON, T., 1895. "The Theory of Light" (second edition) (London : Macmillan and Co.).
STOKES, A. R. and WILSON, A. J. C., 1942. *Proc. Camb. Phil. Soc.*, 38, 313.

THE LIGHT-DIFFUSING PROPERTIES OF MAGNESIUM OXIDE

By V. G. W. HARRISON,

The Printing and Allied Trades Research Association, London

ABSTRACT. Measurements made on magnesium oxide deposited to a depth of 2 mm. on aluminium show that, for angles of incidence up to 60°, the polar reflexion curves are in general *ellipses*, and not the circles required by Lambert's law. For angles of incidence less than 45°, the major axis of the ellipse is perpendicular to the plane of the test surface, while for angles of incidence from 45° to 60° the minor axis of the ellipse is perpendicular to this plane. At 45° incidence, the ellipse degenerates to a circle, and in this special case only, Lambert's law is obeyed to within the accuracy of the measurements. For angles of incidence greater than 60°, the curves become quite irregular, and at almost grazing incidence there is a distinct reddish specular component in the reflected light. An attempt is made to analyse the curves into specular and diffuse components by Barkas's method, but the results suggest that this method is not really applicable in the present instance. 45° appears to be a suitable angle of incidence to adopt in practical gloss determinations.

§1. INTRODUCTION

THOSE interested in measuring the gloss of real surfaces frequently need for reference a reproducible matt surface. The ideal matt surface is unknown; nevertheless there are several conveniently-prepared real surfaces which approximate to it fairly closely under certain conditions of illuminating and viewing; and of these, the surface of magnesium oxide deposited on silver or aluminium, widely used as a standard in colorimetry, merits special consideration. Not only is such a surface easily prepared and of high reflecting power (97–98%), but it is also as nearly a true matt as is any other surface that has so far been suggested.

If, in figure 1, XY represents the ideal matt surface illuminated by a beam IO incident at an angle i, then the intensity of the light scattered in the direction OP is, by definition, proportional to $c . \cos i . \cos R$, where c is a factor defining the brightness of the surface and R is the angle of viewing. More exactly, the *illumination* at the point P per unit area of the diffusing surface is given by the equation

$$L = (cI/d^2) . \cos i . \cos R,$$

where I is the intensity of the incident beam and d is the distance of the point P from O. This relation is the well-known Lambert's law (Lambert, 1760). It is assumed that the dimensions of the diffusing surface are small compared with the distance d. It follows that if I_0 is the intensity of the radiation emitted normally to the surface, then the intensity of the radiation emitted at any other angle R is, for constant conditions of illumination, given by the simple equation

$$I_R = I_0 . \cos R.$$

The apparent simplicity of Lambert's law is deceptive; although quite a number of surfaces obey it approximately under certain conditions of illuminating and viewing, no material is known for which it holds exactly, and none of the several attempts to account for it theoretically has been very successful. The truth of the law is certainly not self-evident, and its value lies only in its convenience and simplicity, since it is the only relation for which the brightness of the radiating or reflecting surface is independent of the angle of viewing. Bouguer (1762) tried to account for Lambert's law on the supposition that the reflecting surface behaves as an aggregate of small mirrors set at all possible inclinations. Lommel (1880) considers the case of glowing masses of metals and gases which emit light not only from their surface, but also throughout their whole volume, and he later attempts to extend this theory to account for the reflecting properties of matt surfaces (1889). Unfortunately the equations derived are of little practical value.
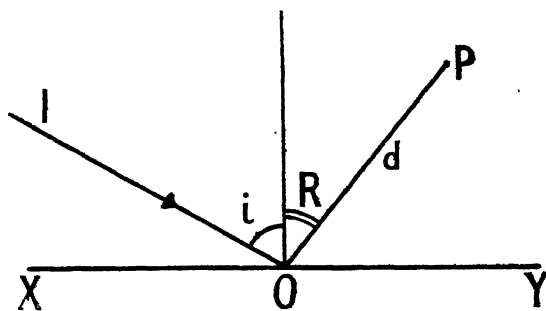
Figure 1.

Further work on this subject has been done by Grabowski (1914), Chinmayanandam (1919), Berry (1923), Pokrowski (1924 to 1926), Schulz (1925) and Barkas (1939). Attempts to discover materials for which Lambert's law is obeyed have been made by Bouguer (1762), Provostaye and Desains (1851), Kononowitsch (1879 and 1881) and Möller (1885). Ångström (1885) found that the polar light-intensity distribution curves were ellipses rather than the circles required by Lambert's law: for the substances examined, he found that the curves were sections of prolate spheroids for small angles of incidence, and oblate spheroids for large angles of incidence; and at an angle of incidence of about 30° the spheroids degenerated into a sphere, so that at this particular angle of incidence the simpler form of Lambert's law is obeyed. Further studies of matt surfaces have been made by Godard (1887), Seeliger (1888), Messerschmidt (1888), Wiener (1892), Jenko (1898), Wright (1900), Thaler (1903), Henning and Heuse (1922) and Woronkoff and Pokrowski (1924),.

§ 2. EXPERIMENTAL DETAILS

(a) *Preparation of specimens*

In order to get a reproducible standard surface for colour and gloss measurements, the magnesium-oxide layer should be infinitely thick. In practice, this is taken to be a layer which is so thick that doubling its thickness produces no appreciable change in its reflecting properties. Unfortunately, magnesium oxide is not as opaque as could be desired, and a deposit of several millimetres must

be made before the layer can be regarded as "infinitely" thick. As the thickness of the layer increases, the tendency for the deposit to fall off its support in large flakes also increases rapidly, and in practice it becomes difficult to work with a layer more than 2 or 3 mm. in thickness; some compromise must therefore be made. For the experiments here described, the diffusing surfaces were prepared by holding a disc of aluminium, 73 mm. in diameter and 5 mm. in thickness, at a distance of about 10 cm. above a piece of burning magnesium ribbon. The process of

Table 1. Effect of the thickness of the magnesium oxide layer on the intensity of the light scattered in various directions (for angle of incidence − 30°)

Angle of viewing (degrees)	Relative intensity for layer of thickness	
	1 mm.	2 mm.
−75	21·4	21·8
−70	29·6	29·8
−65	38·4	38·6
−60	46·1	46·6
−55	54·1	54·9
−50	61·1	62·0
−10	99·2	99·1
− 5	100·0	99·8
0	100·0	100·0
5	99·2	99·0
10	97·8	97·3
15	95·3	95·1
20	92·0	91·9
25	87·9	88·0
30	83·2	83·3
35	77·6	78·9
40	71·3	72·3
45	65·8	66·7
50	59·0	59·7
55	51·8	53·1
60	43·7	44·7
65	36·1	37·2
70	28·0	29·2
75	20·5	21·2

burning the ribbon was continued until the thickness of the deposit reached 2 mm. In order to check whether this thickness was sufficient to reduce the effect of the aluminium support to negligible proportions, goniophotometric measurements were made on layers of 1 mm. and 2 mm. thickness, with the results shown in table 1. It will be seen that, while doubling the thickness of the layer has increased the diffusion slightly (as shown by the slightly greater intensity of light scattered at large angles to the normal), the increase is nevertheless quite small, and a further increase in the thickness of the deposit would not cause appreciable differences in the form of the polar curves.

(b) *Technique of measurement*

All measurements were made on the new Patra Hilger Goniophotometer, an instrument following the general lines of the model described in Patra Research

Report No. 1 (July 1938—now out of print). Particulars of this instrument are to be published shortly. The essential features are given in figure 2. Light from a ribbon filament lamp is collected by a condenser L_1, and the optical system is such that an image of the slit S_1 is thrown on to the test specimen M, the length of the slit being parallel to the axis about which the specimen and photometer arm rotate. The angular spread of the incident beam is controlled mainly by the slit S_2: the smaller the opening, the more nearly parallel is the incident light, but the intensity of the light reaching the photocell P is correspondingly reduced. In practical measurements one has always to effect a compromise between a reasonably small angular aperture and workable galvanometer deflexions. The photocell P is of the ordinary rectifier type connected to a sensitive galvanometer: the angular spread of the wedge of light reaching this receptor is controlled by the slit S_3.

In the arrangement used, the width of the band of light leaving the back surface of the lens L_3 was 12 mm., and the width of the slit image at M was 1 mm. The distance between L_3 and M was 185 mm., so that the total angular spread of the wedge of incident light was $2 \tan^{-1} (55/1850)$, that is, about $3° 24'$. The angular aperture of the receptor system was the same. This aperture would be too great
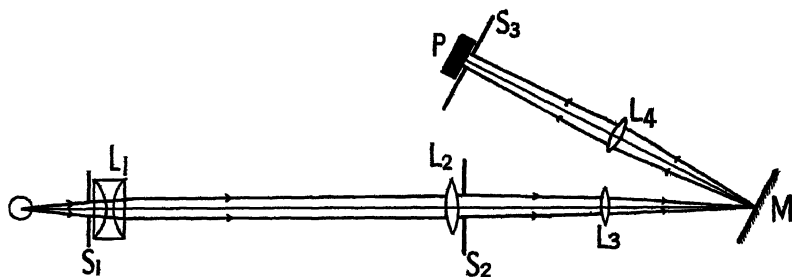
Figure 2.

for materials possessing "peaky" goniophotometric curves and would have to be cut down; but with surfaces of very low gloss it does not introduce serious errors into the form of the curves. More important from the present point of view is the size of the spot of light thrown on to the specimen. It is assumed in all these measurements that the dimensions of the spot of light on M are small compared with the distance of the photocell from the test surface (PM). If this is not so, the curves can be appreciably distorted. In the arrangement used, measurements were made only in the plane containing the angle of incidence, and the length of the spot of light (being at right angles to this plane) was thus of minor importance and was kept to 14 mm. The width of the slit S_1 was reduced until the width of the rectangular spot of light thrown on to M was 1 mm. for normal incidence (for oblique angles of incidence the width of the slit image becomes correspondingly greater). The distance from L_4 to M in the instrument used was 17 cm. and from L_4 to P, 20 cm. The focal length of the lens L_4 was about $18\frac{1}{4}$ cm.

To show the effect of the size of the spot of light on the form of the curves, the figures in table 2 are reproduced. Those in columns 2 and 3 were taken with the goniophotometer, now destroyed, described in Patra Research Report No. 1. This instrument used a photocell of about $3\frac{3}{4}$ cm. diameter placed at a distance of 50 cm. from the test surface; no lens was used in the receptor system. A circular

spot of light was thrown on to the specimen (that is, at normal incidence, the circle being broadened into an ellipse for oblique incidence), and this circle was usually of 1 cm. diameter: the figures given by this arrangement are in column 2. The corresponding figures given when the spot of light was cut down to 4 mm. diameter are in column 3. The effect on the curves is clearly seen. Finally, in column 4

Table 2. Effect of the dimensions of the spot of light thrown on to the
test specimen (for angle of incidence −45°)

Angle of viewing (degrees)	Apparent relative intensity for spot of light		
	10 mm. diam. circle	4 mm. diam. circle	14 × 1 mm. rectangle
−75	25·5	22·7	22·9
−70	34·8	31·4	31·4
−65	43·8	40·0	40·2
−25	93·0	91·3	90·9
−20	95·4	94·7	94·9
−15	97·5	97·1	97·1
−10	98·8	99·0	99·3
− 5	99·5	99·8	100·0
0	100·0	100·0	100·0
5	99·5	99·8	99·3
10	98·8	98·8	98·0
15	97·5	97·0	95·1
20	95·7	94·7	93·4
25	93·1	91·8	90·0
30	89·9	88·4	86·0
35	86·1	84·2	81·8
40	82·1	79·4	76·0
45	77·2	74·0	70·1
50	71·9	68·2	63·8
55	65·3	61·8	57·1
60	58·3	54·8	49·7
65	50·8	47·3	41·9
70	42·1	39·2	34·1
75	32·7	30·6	25·0

are the figures obtained with the present instrument, with the rectangular spot of light 14 × 1 mm. The figures are all corrected so as to give a reading of 100·0 when the angle of viewing is 0°. The broadening of the curves with large spots of light is attributed to the fact that with a large spot of light the effective area sending flux to the photocell tends to increase as the angle of viewing increases.

§3. RESULTS

The results of the measurements on magnesium oxide are given in table 3, which shows the measured intensity for various angles of incidence and viewing, expressed as a percentage of the maximum intensity obtainable with the given settings of the instrument. This maximum intensity occurs when the directions of incidence and viewing are both normal to the test surface. The intensity cannot be measured experimentally under these conditions, but it can be determined to a satisfactory degree of accuracy by extrapolation, since it is found that

Table 3. Relative intensity of the light scattered from a surface of magnesium oxide at various angles of incidence and viewing

Angle of viewing	Angle of incidence															
	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
−75	19·8	20·1	20·4	20·1	20·4	20·4	20·4	20·2	20·7	21·1	21·7	23·1	—	—	—	—
−70	27·9	27·2	27·6	27·6	27·2	27·6	27·9	27·9	28·2	28·9	28·5	—	—	—	—	—
−65	35·0	34·6	35·6	35·3	35·0	35·3	35·6	35·7	36·0	37·0	—	—	—	—	—	—
−60	42·1	43·1	43·1	43·1	42·7	44·1	43·5	43·4	43·8	—	—	—	—	—	—	48·1
−55	49·8	50·9	51·2	50·6	50·6	52·2	51·6	51·2	—	—	—	—	—	—	53·2	52·6
−50	57·1	58·0	58·8	58·3	57·8	59·0	59·1	—	—	—	—	—	—	—	57·7	56·8
−45	63·6	65·5	65·5	64·8	65·5	66·1	—	—	—	—	—	—	—	61·0	61·9	60·8
−40	70·8	71·8	72·7	71·9	72·0	—	—	—	—	—	—	—	66·8	64·8	65·6	64·2
−35	76·1	77·3	79·1	77·8	—	—	—	—	—	—	—	73·2	70·8	69·4	68·8	66·8
−30	81·7	83·0	84·2	—	—	—	—	—	—	—	78·7	77·1	74·7	72·8	71·8	69·4
−25	86·5	87·7	—	—	—	—	—	—	—	83·8	82·3	81·1	77·3	75·8	74·7	71·4
−20	91·4	—	—	—	—	—	—	—	88·7	87·4	85·0	84·8	80·1	77·8	76·0	73·2
−15	—	—	—	—	—	—	—	92·1	91·4	89·7	87·7	86·7	82·3	80·1	77·6	74·7
−10	—	—	—	—	—	—	94·1	94·0	92·8	91·6	89·7	88·2	83·8	81·8	78·6	75·8
−5	—	—	—	—	—	96·5	94·7	94·7	93·9	92·2	90·9	88·7	84·8	82·2	—	75·9
0	(100)	—	—	—	96·7	96·7	95·0	94·7	94·1	92·2	91·2	88·7	85·1	82·8	79·2	75·9
5	—	—	—	96·7	96·0	96·0	94·2	94·7	93·5	91·6	90·2	88·2	85·1	82·4	79·2	75·9
10	—	—	97·2	95·2	94·7	94·7	93·0	93·0	92·1	90·1	88·8	87·0	84·0	82·2	78·9	76·1
15	—	94·0	95·2	92·7	92·1	92·1	90·9	90·8	89·7	87·3	86·7	85·7	82·3	81·1	78·2	75·2
20	92·1	90·3	92·1	88·7	89·5	89·7	88·2	86·7	86·2	86·1	84·8	83·7	81·8	79·8	77·2	74·9
25	87·2	86·2	86·9	84·2	85·7	85·7	84·3	83·1	82·9	83·1	81·8	81·1	79·6	77·9	75·8	73·3
30	82·2	81·1	82·2	79·8	80·8	81·1	80·4	79·2	79·2	79·1	78·5	78·2	77·1	75·9	74·7	71·8
35	76·6	75·8	76·5	74·6	75·2	76·1	75·2	74·7	74·7	75·2	74·7	74·4	73·9	73·3	72·2	70·8
40	70·4	69·7	70·0	68·8	70·0	70·8	70·1	69·3	69·7	70·1	70·1	70·1	70·1	70·8	70·4	70·8
45	63·6	63·2	63·9	62·6	63·6	64·2	63·9	64·1	63·9	64·6	65·8	66·2	66·2	69·7	68·2	68·8
50	56·3	55·8	57·0	55·1	56·3	57·3	57·3	58·0	58·1	58·7	60·3	61·7	62·3	64·6	65·7	67·1
55	48·6	48·6	49·5	48·0	49·3	49·9	50·7	51·2	51·7	52·6	54·4	56·1	57·0	61·0	62·7	66·2
60	40·8	40·9	41·5	41·5	41·5	42·5	43·1	44·4	44·4	45·7	48·0	50·3	52·7	56·4	59·8	64·2
65	33·1	32·7	33·7	33·7	33·7	35·0	35·6	36·4	37·3	38·6	41·2	43·5	46·4	51·7	57·1	62·2
70	24·9	25·3	25·9	25·9	26·3	27·3	27·9	28·9	29·2	31·4	33·4	36·0	39·2	45·1	52·9	62·8
75	17·5	17·8	18·2	17·8	18·2	19·1	20·1	20·7	21·4	23·0	25·0	27·9	30·8	36·4	44·1	60·3

the experimental curves are very nearly ellipses. The figures are in fairly good agreement with those of Barkas (1939).

Examination of these results shows at once that the variation with i predicted by Lambert's law is not exactly true. If it were, the *flux* reaching the photocell would be independent of the angle of incidence. We find, in fact, that the flux received by the photocell is only approximately constant and independent of the angle of incidence when the angle of viewing is at the special angle of 40°. The simpler form of Lambert's law ($I_R = I_0 \cdot \cos R$) holds reasonably well for an angle of incidence of 45°, but it does not hold accurately at any other angle of incidence.

The polar curves for angles of incidence less than 45° are found to be ellipses with the major axis perpendicular to the test surface. The polar equation of an ellipse with origin of polar co-ordinates at one end of the major axis is

$$I = \frac{2ab^2 \cdot \cos R}{a^2 \sin^2 R + b^2 \cos^2 R},$$

where a and b are the semi-major and semi-minor axes. By a suitable choice of a and b, this equation expresses the experimental results to within the accuracy
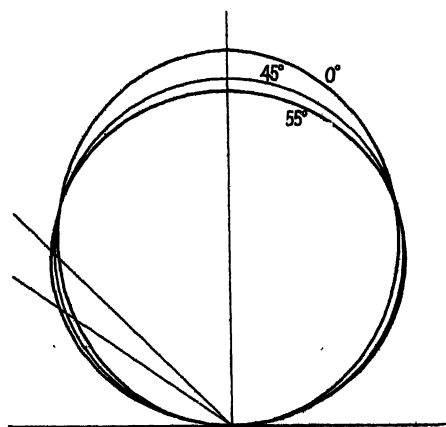
Figure 3.

of measurement. Table 4 shows the measured and calculated values for angles of incidence of 0°, 15°, 25°, 30°, 40°, 45°, 50°, 55° and 60°. It will be seen that in most cases the agreement is good. For angles of incidence less than 45° the curves thus represent sections of a prolate spheroid, as recorded by Ångström (1885). At 45° incidence the ellipse degenerates to a circle. For angles of incidence between 45° and 55° the curves are approximately sections of an oblate spheroid—that is, they are again ellipses, but this time with the *minor* axis perpendicular to the plane of the surface. At 60° incidence, the ellipse begins to lose its symmetry and develops a " bulge " in the region of 60° to 70°; for angles of incidence of 65° and over, the curves become quite irregular and no simple equation can be found to fit them; and at nearly grazing incidence (80° and over) there is distinct regular reflexion from the surface, which, at the appropriate angle of viewing, looks polished with a reddish tinge. The polar curves for 0°, 45° and 55° incidence are drawn in figure 3.

Table 4. Comparison of observed values of intensity with those calculated
from the ellipse $I = \dfrac{2ab^2 \cos R}{a^2 \sin^2 R + b^2 \cos^2 R}$

	Angle of viewing									
	0	10	20	30	40	50	60	70	80	90
$i=0°$, $a=50.0$, $b=45.0$										
I, calculated	100	97.8	92.4	82.0	69.7	56.5	42.7	28.4	14.2	0
I, observed			92.1	82.2	70.4	56.3	40.8	24.9		
			91.4	81.7	70.8	57.1	42.1	27.9		
$i=15°$, $a=48.7$, $b=44.5$										
I, calculated	97.4	95.5	90.0	80.2	68.9	56.1	42.4	28.4	14.2	0
I, observed		95.2	88.7	79.8	68.8	55.1	41.5	25.9		
					71.9	58.3	43.1	27.6		
$i=25°$, $a=48.3$, $b=45.4$										
I, calculated	96.7	94.9	90.0	80.9	69.9	57.2	43.8	29.5	9.9	0
I, observed	96.7	94.7	89.7	81.1	70.8	57.3	42.5	27.3		
						59.0	44.1	27.6		
$i=30°$, $a=47.5$, $b=44.6$										
I, calculated	95.0	93.1	88.1	79.3	69.0	57.0	43.4	29.3	14.7	0
I, observed	95.0	93.0	88.2	80.4	70.1	57.3	43.1	27.9		
		94.1				59.1	43.5	27.9		
$i=40°$, $a=47.0$, $b=45.5$										
I, calculated	94.2	92.3	88.0	80.0	69.8	58.2	44.5	30.3	15.5	0
I, observed	94.1	92.1	86.2	79.2	69.7	58.1	44.0	29.2		
		92.8	88.7				43.8	28.2		
$i=45°$, $a=46.1$, $b=46.1$										
I, calculated	92.2	90.8	86.6	79.9	70.6	59.2	46.1	31.6	16.0	0
I, observed	92.2	90.1	86.1	79.1	70.1	58.7	45.7	31.4		
		91.6	87.4					28.9		
$i=50°$, $a=45.6$, $b=46.0$										
I, calculated	91.2	89.8	86.2	79.2	70.4	59.4	46.1	31.8	16.1	0
I, observed	91.2	88.8	84.8	78.5	70.1	60.3	48.0	33.4		
		89.7	85.0	78.7				28.5		
$i=55°$, $a=44.4$, $b=46.8$										
I, calculated	88.7	87.9	84.6	78.9	71.0	60.9	48.0	33.4	17.2	0
I, observed	88.7	87.0	83.7	78.2	70.1	61.7	50.3	36.0		
		88.2	84.8	77.1						
$i=60°$, $a=42.6$, $b=47.5$										
I, calculated	85.4	84.4	82.0	77.2	70.8	62.0	49.8	35.4	18.4	0
I, observed	85.1	84.0	81.8	77.1	70.1	62.3	52.7	39.2		
			83.8	80.1	74.7	66.8				

No theoretical explanation is here advanced as to why the polar curves should be ellipses rather than circles. These results are presented merely as experimental facts. Some other expressions were examined to see whether the

experimental results would fit them better than the equation of the ellipse. Thus the equation

$$I = I_0 \cos R \cdot e^{-Kt. \sec R}$$

holds approximately for the light transmitted through certain diffusing layers (Harrison, 1940), where t is the thickness of the layer and K is a constant. The equation gives families of oval curves. No values of Kt could be found which gave better agreement between observed and calculated values than is shown in table 4. The equations of Henning and Heuse (1922), Lommel (1880, 1889) and Ångström (1885) were also examined, but no real values of the constants could be found such that these equations would fit the observed results.

§4. THE BARKAS ANALYSIS

The results were examined by the Barkas analysis (1939), in which the reflecting properties of the surface are expressed in terms of an optically equivalent surface made up of mirror facets and rough facets, tilted at all angles. The fundamental equation may be written (in our present notation) as

$$I_{i,R}, \cdot \cos i = K\{\cos (i - R) + \cos (i + R)\} + sB \cdot \cos \tfrac{1}{2}(i - R),$$

where K is a factor defining the contribution of the rough elements, while B is a similar factor defining the contribution of the mirror elements. s is a function

Table 5. Values of K [Barkas's $\tfrac{1}{2}(rA)$] for increasing values of $(i+R)$

$(i+R)$	K
0	39·8
5	40·4
10	40·1
15	41·3
20	42·9
25	42·4
30	42·4
35	42·8
40	41·5
45	41·8
50	40·7
55	41·2
60	38·8
65	36·1
70	(37·6)
75	(33·4)

of $\tfrac{1}{2}(i - R)$ and is calculated from Fresnel's laws of reflexion. Both K and B are in general functions of i and R.

In accordance with Barkas's treatment, $I_{i,R} \cdot \cos i$ was plotted against $(i - R)$, the curves obtained being closely similar to, but not identical with, those published by Barkas. The values of K [Barkas's $\tfrac{1}{2}(rA)$] were calculated from the curves by his method over the range $(i+R) = 0°$ to $75°$, with the results shown in table 5. The last two values (in parentheses) cannot be regarded as very accurate, and it is not possible to determine K for values of $(i+R)$ greater than $75°$ by Barkas's method:

there is, however, clear evidence that the value of K is beginning to fall fairly rapidly for values of $(i+R)$ greater than 50°. Barkas quotes a value of 0.41_5 for K for all values of $(i+R)$ up to 70°, and he assumes that this value of K holds good when $(i+R)$ is greater than 70° (when K cannot be determined directly). Examination of table 5 will show that, while the mean value of K over the whole range (40·4) is in satisfactory agreement with Barkas's value, nevertheless the figures do show a decided trend, and suggest that K rises to a maximum at about $(i+R) = 35°$, and thereafter diminishes with increasing rapidity.

Table 6. Values of B for $K = 0.41$ (assumed constant)

$(i-R)$	$(i+R)$				
	0	10	20	30	40
0	2·5	2·2	1·9	1·5	0·9
10	2·2	2·0	1·7	1·3	0·7
20	1·7	1·6	1·2	0·8	0·3
30	1·2	1·2	0·6	0·0	—0·1
40	0·9	0·9	0·3	—0·2	—0·2
50	0·8	0·6	0·1	—0·3	—
60	0·9	0·8	0·2	—	—
70	1·0	—	—	—	—

There are further difficulties. If, despite the evidence, one assumes that K is, in fact, constant over the whole range, and one calculates the values of B according to the method which Barkas has outlined in his paper, one gets the results in table 6. These values are in satisfactory agreement with those given by Barkas, but there are some negative values. This does not seem plausible; moreover, the table suggests that in the cases where it has not been
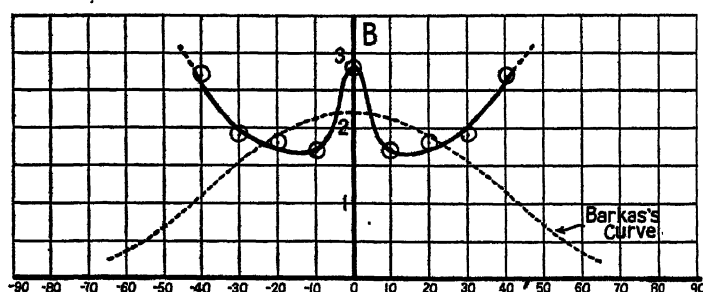
Figure 4.

possible to determine B experimentally, it may have even greater negative values than are here recorded. On the other hand, if we now, instead of assuming that K is constant, take the experimental values given in table 5, we get the values of B given in table 7. Here the values of B do not agree with those of Barkas, and there is a marked tendency for them to rise again after passing through a minimum near $(i+R) = 20°$. When plotted, they give the curve shown in figure 4, which expresses the comparative areas of the mirror facets at various

inclinations to the mean surface: the form of this curve is a good deal less plausible than the one given by Barkas (shown dotted), though one can hardly say definitely that it is wrong. The truth seems to be that the Barkas analysis

Table 7.　Values of B for K varying between
42·9 and 33·4, according to table 5

$(i-R)$	$(i+R)$				
	0	10	20	30	40
0	2·8	1·7	1·8	1·9	2·7
10	2·5	1·5	1·6	1·7	2·5
20	2·0	1·1	1·1	1·2	1·8
30	1·5	0·8	0·5	0·6	1·0
40	1·4	0·2	0·2	0·0	0·2
50	1·0	0·3	0·0	0·0	0·0
60	1·0	0·5	0·2	—	—
70	1·2	—	—	—	—

is not very satisfactory when applied to surfaces as matt as magnesium oxide, and unless K can be determined accurately for a much wider range of values of $(i+R)$ the values of B will remain open to considerable doubt.

§ 5. SUMMARY AND CONCLUSIONS

It has been shown that Lambert's law does not hold for a surface of magnesium oxide 2 mm. thick, since the flux emitted per unit area is not proportional to the cosine of the angle of incidence. The simpler form of the law,

$$I_R = I_0 \cdot \cos R,$$

is valid only in the special case where the angle of incidence is 45°. The polar curves for angles of incidence below 60° are represented to within the accuracy of measurement by ellipses, the general equation of which is

$$I = \frac{2ab^2 \cdot \cos R}{a^2 \sin^2 R + b^2 \cos^2 R}.$$

Above 60° incidence the curves become decidedly irregular and cannot be expressed in terms of any simple equation. Above 75° incidence signs of specular reflexion appear, and at nearly grazing incidence there is a fairly well defined reddish specular component.

This is not the place to discuss the design of glossmeters. In practice, one finds instruments making use of angles of incidence from 45° to over 80°. Never-theless, the fact that 45° incidence is the only angle for which Lambert's law in its simpler form holds for magnesium oxide seems to be a fairly powerful argument for the adoption of 45° incidence in practical work, particularly as 45° is a very convenient angle from the point of view of instrument design. The point is, with magnesium oxide at 45° incidence we have a conveniently prepared surface which does give a polar curve approximating closely to Lambert's circle, while at 60° incidence and more, it is doubtful whether there is any real

surface for which Lambert's law is even approximately true. At 45° incidence, therefore, we have a useful reference material of zero gloss, while at 60° incidence and over, we have none.

ACKNOWLEDGMENT

Finally, my thanks are due to the Council of the Printing and Allied Trades Research Association for permission to publish this paper.

REFERENCES

ÅNGSTRÖM, K., 1885. *Ann. Phys., Lpz.*, **26**, 253.
BARKAS, W. W., 1939. *Proc. Phys. Soc.*, **51**, 274.
BERRY, E. M., 1923. *J. Opt. Soc. Amer.*, **7**, 627.
BOUGUER, P., 1762. *Traité d'optique* (Richtenburg's Latin translation, Book 2, Sect. 3, Art. 3).
CHINMAYANANDAM, T. K., 1919. *Phys. Rev.*, **13**, 96.
GODARD, L., 1887. *Ann. Chim. (Phys.)*, **10**, 354.
GRABOWSKI, L., 1914. *Astrophys. J.*, **39**, 299.
HARRISON, V. G. W., 1940. *Proc. Tech. Sect. Paper Makers' Assoc.*, **21**, 67.
HENNING, F. and HEUSE, W., 1922. *Z. Phy..*, **10**, 111.
JENKO, P., 1898. *Ann. Phys., Lpz.*, **66**, 1182.
KONONOWITSCH, A., 1879. *Fortschr. Phy..*, **35**, 430 ; 1881. *Ibid.*, **37** (2), 481.
LAMBERT, J. H., 1760. *Photometria* (Augsburg), Part 1, Cap. 2, Sections 53, 72, 77–81.
LOMMEL, E., 1880. *Ann. Phys., Lpz.*, **10**, 449 ; 1889. *Ibid.*, **36**, 473.
MESSERSCHMIDT, J. B., 1888. *Ann. Phys. Lpz.*, **34**, 867.
MÖLLER, W., 1885. *Ann. Phys., Lpz.*, **24**, 266.
POKROWSKI, G. I., 1924. *Z. Phys.*, **30**, 66 ; 1925. *Ibid.*, **35**, 34 ; 1926. *Ibid.*, **35**, 390 and 472.
DE LA PROVOSTAYE, F. and DESAINS, P., 1851. *C.R. Acad. Sci.*, Paris, **33**, 444. Also 1847. *Ibid.*, **24**, 60 ; *Ibid.*, 1848. **26**, 212 ; 1852. *Ann. Chim. (Phys.)*, **34**, 192.
SCHULZ, H., 1925. *Z. Phys.*, **31**, 496.
SEELIGER, H., 1888. *Sitzungsberichte der math. phys. Klasse der Königl. bayerischen Akad. der Wissenschaften zu München*, **18**, 201.
THALER, F., 1903. *Ann. Phys., Lpz.* (4), **11**, 996.
WIENER, C., 1892. *Ann. Phy..., Lpz.*, **47**, 638.
WORONKOFF, G. P. and POKROWSKI, G. I., 1924. *Z. Phys.*, **20**, 358.
WRIGHT, H. R., 1900. *Phil. Mag.*, **49**, 199.

THE BEHAVIOUR OF WATER UNDER HYDROSTATIC TENSION: I

By H. N. V. TEMPERLEY and LL. G. CHAMBERS,

Admiralty Research Laboratories

ABSTRACT. A study is made of the behaviour of water under hydrostatic tension. A review of the literature reveals very considerable differences in the values of critical tension obtained by different methods, the factor of disagreement being sometimes as high as 50. Two methods are studied in detail: the Berthelot method, which makes use of the difference in expansions of glass and water and the Reynolds centrifugal method. Evidence is brought forward that it requires high pressures to persuade water to stick to glass under subsequent tension. A modification of the Berthelot method gives results in much better agreement with other methods. Examination of the Reynolds method shows that account must be taken of the fact that water does not move as a rigid body. Although this has no appreciable effect on the pressure distribution, it does imply a considerable stirring of the water, and any small bubbles would tend to be brought near the region of greatest tension in a time that might be comparable with that of the experiment.

The flow of water in constricted tubes is also studied. The critical tension appears to be low, but this is to be expected in view of the fact that the motion is probably turbulent before cavitation occurs.

It is concluded that if tension is applied statically, ordinary water can stand tensions of the order of 40 atmospheres, even if it is not perfectly air-free. Water nearly saturated with air has been shown to stand tension up to 6 atmospheres.

§ 1. INTRODUCTION

IT has long been known that, under suitable conditions, liquids are capable of sustaining considerable tensions. While under tension, the liquid is in a metastable state and, if the tension is too great, it changes irreversibly to the two-phase system liquid and vapor and dissolved gas. The critical tension is of interest in quite a number of fields of pure and applied science, and experiments have been made by a large number of workers. As the published results are very discordant, an attempt has been made to clear up the causes of some at least of the disagreements between various methods. The following are some of the fields of research in which the problem occurs:

(*a*) The ascent of sap in plants, and the ejection of spores from ferns.

(*b*) The formation of bubbles in liquids, e.g. blood supersaturated with gas.

(*c*) The design of ships' propellors.

(*d*) The design of echo-sounding and similar equipment, and the study of the biological effects of supersonic waves.

(*e*) The study of surface phenomena attending an under-water explosion.

It is believed (Dixon and Joly, 1895) that the ascent of sap is due, at least in part, to actual hydrostatic tension which is supplied by capillarity in the passages of the leaf. The ejection of spores (King, 1944) is sometimes achieved by a catapult-like mechanism depending on the rupture of a gradually-drying drop of

water.   Near ships' propellers, negative pressures occur due to the inertia of the water; near sources of high-frequency sound, the amplitude may exceed the hydrostatic pressure so that the water is in tension for part of the cycle.   Negative pressures are not the direct result of an under-water explosion, but occur when the resulting pressure-pulse is reflected at a yielding surface such as the free surface of water.   In order to keep the pressure at such a surface zero or small, the pressure-pulse is reflected in part as a tension-pulse, so that all the water between the charge and the free surface is subjected first to pressure, then to tension.

§ 2. SUMMARY OF RESULTS OBTAINED BY OTHER WORKERS

Useful reviews of the literature have been given by Harvey and others (1944) and by Vincent (1941 ; 1943 a and b), who have themselves made valuable contri-butions to the subject.   For convenience we subdivide the results into three classes, according to the tensions estimated.   (We have been unable to examine all the papers ourselves.)

Class A.   Maximum tensions over 100 atmospheres

The chief work in this class is that of Dixon (1909).   His method was that introduced by Berthelot of heating liquid in a sealed tube until it filled, then cooling the whole until the liquid broke.   Values of tension were then estimated from the corresponding temperature difference, assuming values for compressibility of water and coefficient of expansion determined at zero pressure, due corrections being made for the thermal expansion of the glass and its elastic deformation. He estimated tensions up to 200 atmospheres.

To this work should perhaps be added that of Kenrick, Wismer and Wyatt (1924), who saturated liquids with gas at 100 atmopheres and reduced the pressure to 1 atmosphere without effervescence, and that of Kenrick, Gilbert and Wismer (1924), who heated water to 270° c. under atmospheric pressure before it exploded. A liquid superheated or supersaturated with gas stands in much the same relation to a liquid under hydrostatic tension as does a solid vessel stressed by pumping air into it to the same vessel stressed by evacuating the region outside.   The presence or absence of gas is unlikely to affect the tensile strength of a solid, but might well affect that of a liquid, and might alter the " microscopic " distribution of stress.

Harvey (1944) found that water previously compressed to 1000 atmospheres can withstand tensions (occurring as the result of a hammer blow) estimated at from 100 to 1000 atmospheres, but no definite figures are available.   King (1944), from a study of the catapult-like mechanism of the projection of spores in the common fern, arrives at a value of 200 atmospheres for the critical strength of water in the presence of plant materials, but his data would be consistent with much lower or higher values.

Class B.   Maximum tensions between 10 and 100 atmospheres

Near the top of this class stands the work of Budgett (1912), who measured the force necessary to pull apart slightly wet optically flat steel surfaces which had been " wrung " together.   On the crude assumption that the liquid film is continuous over the surface, one deduces a value of only 4 atmospheres for the critical tension, but Budgett, from examination of the surfaces after separation, and by comparison

of the results under vacuum with those under atmospheric pressure, estimated that air could reach most of the inner surfaces and that the effective area of the water was only about 7% of that of the steel. On this basis he estimated that the critical tension water could stand was about 60 atmospheres. He gave arguments which suggest that only a little of this tension is due to capillarity, and also that the failure occurs in the liquid rather than at the steel surfaces. Perfectly dry surfaces did not adhere, so the attractions of the steel surfaces contributed nothing to the effect. ·

Next come a number of measurements by Worthington (1892) and Meyer (1911). Their method is similar in principle to that of Berthelot, that is, the tension is produced by cooling liquid in a glass vessel, but here the tension is actually measured instead of being inferred from the temperature difference. Worthington measured the tension by noting the change of volume of an ellipsoidal bulb containing mercury (and enclosed in the apparatus), Meyer by means of the rotation of a glass spiral containing the liquid. These devices had to be calibrated with positive pressure and extrapolated to the region of tension. Meyer obtained a value of 34 atmospheres for water and Worthington one of 17 atmospheres for ethyl alcohol.

Yet another modification of Berthelot's method was introduced by Vincent (1943 b). In order to make sure that the pressure in the tube was truly zero at the moment of filling, the " sealing " was carried out by freezing part of the liquid in a capillary tube attached to the main tube and open to the atmosphere at the far end. Unfortunately the liquid used was mineral oil and not water, but the mineral oil had been previously tested by the ordinary Berthelot method and had given a value of 119 atmospheres (maximum) for the tension compared with 157 atmospheres for water under similar conditions. By the modified Berthelot method, in which the pressure at the instant of sealing was definitely known to be atmospheric, values for the critical tension ranging from zero up to 25 atmospheres were obtained. In a further set of readings, where the tube was sealed, the liquid heated to generate pressure, the tube unsealed and the pressure adjusted, and the tube sealed again and cooled in order to apply tension, values for the critical tension varying from 0·9 to 24·5 atmospheres were obtained, which appeared to indicate that the pre-application of pressures of the order of 10 to 100 atmospheres did not alter the results in order of magnitude.

Momentary tensions of the order of 15 atmospheres have actually been measured by piezo-electric gauges in the water near an explosion.

Class C. Maximum tensions below 10 atmospheres

Near the top of this class comes the work of Vincent (1943 a). His method consisted of cooling liquid in a vessel, thus sucking it backwards through a capillary tube. (He calls this the "viscosity tonometer" method). By comparison of the rates of flow just before and just after the break one can infer the critical tension from the ratio of these rates of flow, because the pressure in a bubble in the liquid must be just its vapour pressure. Provided the flow is not turbulent, the rate of flow is proportional to the pressure difference. Vincent obtained a maximum value of 7·8 atmospheres for the same oil that he investigated by the Berthelot and modified Berthelot methods. (This method is unsuitable for water because of its low viscosity.) A method of applying tension to a liquid by centrifugal force was

introduced by Reynolds (1878), who rotated a sealed U-tube on a lathe, one arm of the tube being full of water and the other containing water and water-vapour. He obtained a value of 4·8 atmospheres for water, and Worthington (1892) applied the same method to other liquids, obtaining a value of 7·9 atmospheres for ethyl alcohol.

Dixon and Joly (1895) obtained a few results, the largest being 7·5 atmospheres, by experiments similar to those of Dixon (1909), but the volume of the cavity formed by contraction of the liquid after failure was obtained by measurement of its geometrical dimensions rather than by temperature difference.

In another series of experiments, Vincent (1941) subjected liquid to tension by the very direct method of enclosing it within a metal bellows and applying tension to the bellows. He obtained maximum values of tension of 2·38 atmospheres for ethyl alcohol and 2·94 atmospheres for the mineral oil.

Values of tension for which cavitation occurs near ship's propellers do not seem to have been measured, and data on quartz or magnetostriction oscillators are hard to find in the literature. Dean (1944) quotes one case where cavitation occurred near an oscillator even though the amplitude of the oscillations of pressure at the surface was only 0·8 atmosphere, and the pressure therefore always positive, according to classical hydrodynamics. He suggested that there might be a further lowering of pressure due to rotational motion. According to the data given by Kornfeld and Suvarov (1944), the amplitude of pressure variation when strong cavitation set in would come out classically at just over 1 atmosphere, but again one cannot exclude the possibility of vortices occurring. It seems unlikely that the tensions are very large.

§3. DISCUSSION OF THESE RESULTS

A system consisting of liquid under tension is unstable with respect to a mixture of the liquid and its vapour, so a certain amount of spread in the observed values of tension would not be at all surprising, but the observed spread is enormous. For example, the same observer (Vincent), working with the same liquid (mineral oil), obtained four different values by four different methods, ranging from 2·9 to 119 atmospheres, and a spread of similar extent is found by comparing the results for water. In addition, all the observed results are enormously below the theoretical values. For water the cohesion, or "intrinsic pressure", calculated on the Laplace theory of capillarity, comes out at about 10,000 atmospheres. More refined considerations, taking account of the "hole" structure of the liquid explicitly (Fürth, 1941), still lead to a value of about 3000 atmospheres. A similar discrepancy between the observed and calculated strengths of solids has long been known. Joffé (1928) has shown that part at least of the latter discrepancy is due to surface cracks leading to a non-uniform stressing. Such explanations are not possible for liquids, since stresses cannot be "locked up" in part of a liquid, but must distribute themselves uniformly through it.

It seems certain that the assumptions underlying some of the methods are faulty and that not all of the observers are really measuring the same thing. For example, the liquid may really be failing by tearing away from the supporting surface of the containing vessel so that the properties of this are important. (This certainly seems to occur with the Berthelot method.) This might help to explain

some of the discrepancies between the various methods, but hardly that between theory and experiment. (If it was a fact that the cohesion was immensely stronger than the adhesion, the liquid would not wet the surface at all.) Again, it is quite likely that the critical tension in a liquid may depend on whether the stress is applied statically or dynamically. Such effects are familiar in solids. Yet a third possibility is that a prior application of pressure may alter the properties of the liquid under tension or may make it adhere better. Harvey's work (1944) seems to point to an effect of this kind.

We shall now enumerate and discuss some assumptions made by the observers. There is first the assumption that certain physical quantities measured at zero or positive pressure can be assumed to retain the same, or nearly the same, values under conditions of tension. In Dixon's method (1909) it is assumed that the compressibility and coefficient of expansion of water are unaffected in order of magnitude by tension. In Worthington's (1892) and Meyer's (1911) method it is assumed that a calibration of their measuring instruments for positive pressures holds good for tensions. In Vincent's tonometer method (1943 a) it is assumed that the viscosity is unaffected by tension. It seems most unlikely that Worthington's and Meyer's assumption that the changes of shape of a glass vessel are equal and opposite for pressure and tension can be seriously in error, because glass obeys Hooke's law until near its breaking point. There is rather more doubt about the physical properties of liquids under tension, because they have not been measured. If it were possible to apply sufficient tension, the properties of the liquids would certainly be changed, just as they are at high pressures, but it seems certain that these tensions would have to be comparable with the *theoretical* breaking strength—that is, with the energy necessary to remove the molecules from one another's influence, which quantity can be inferred from the latent heat or surface tension. The only experimental results that both involve this assumption and point to tensions at all comparable with the theoretical strength seem to be those of Dixon (1909). Any criticism of Dixon's work on these lines, implying that the extensibility of water may be much greater than the compressibility, and his inferred tensions therefore too high, tends to stultify itself, because it implies that the actual values of tension are lower, and thus lessens the probable departure of the properties of water from those measured at zero pressure.

A second assumption implicit in Dixon's work is that the pressure in the tube is nearly zero at the moment when it fills with liquid. Vincent and Simmonds (1943) have brought forward some evidence that this assumption is in error for his mineral oil, and later in this paper we shall bring forward other evidence for water (also pointing in this direction). An assumption of a different kind is implicit in the Reynolds centrifugal method. One is inclined rather hastily to assume that the liquid in the tube rotates like a rigid body and that the liquid is being stressed statically. Later in this report we shall examine the situation in detail, when it will appear that the stress is dynamic rather than static, and that conditions are in another respect unfavourable for the development of large tensions, so that the centrifugal method cannot fairly be compared with the truly static methods. In addition, it is almost impossible to avoid vibration completely. The resulting accelerations mean that the *maximum* tension actually applied to the water is greater than the "centrifugal" value.

We feel bound to take exception to one statement by Vincent (1941, p. 128): "The only two methods that are theoretically irreproachable, those of Osborne Reynolds and Budgett, give comparatively low results". We have pointed out above that Budgett's experiments give a rather high result if his own interpretation of them is correct, and we see no reason to question it. Secondly, we hope to show below that Reynolds' method does not measure precisely the same physical quantity as do static methods, so that at least in this sense it is not "theoretically irreproachable".

§4. DESCRIPTION OF WORK CARRIED OUT AND OBJECTIVES AIMED AT

The objectives of the work now to be described were threefold.

(a) To try and find some reasons for the wide discordance of observed results.

(b) To ascertain something about the probable behaviour of water under impulsive, as distinct from static, loading.

(c) To investigate possible reasons for the disagreement between theory and experiment.

It is believed that objective (a) has been partly attained in that acceptable explanations have been found for the very high results obtained by Dixon using the Berthelot method and the low ones obtained by the centrifugal method (Reynolds, 1878) and by Vincent with the metal bellows. Objective (b) requires further work, but the indications are that water is weaker dynamically than it is statically. Harvey's work should help progress towards objective (c). He found that strong centrifuging or pre-compression of water to 1000 atmospheres materially increased its power to withstand tension, probably by the removal of stray nuclei and dissolved gases. The work carried out may be subdivided into three parts which will be described separately. The first part was an examination of the Berthelot method of measuring tension. A modification was made which is believed to be capable of giving more reliable results. The second part was an examination of the Reynolds method. The third part was a brief study of cavitation occurring when water is sucked through a constricted tube.

§5. EXAMINATION OF THE BERTHELOT METHOD OF MEASURING TENSION

Suspicion about this method was first aroused by the fact that attempts to repeat Dixon's experiments in wider tubes (about 5 mm. bore), invariably led to the tube bursting, before it filled, however slowly it was heated. A slightly smaller and stronger tube of 3 mm. bore and 1 mm. wall was then used. These experiments were successful but a large percentage of bursts still occurred. (The bursting strength of such tubing is given by Kaye and Laby as 280 atmospheres, but they recommend a large factor of safety to allow for "locked-in" stress). Attempts were made to repeat the work of Dixon and Joly (1895) by measuring geometrically the size of the cavity formed by shaking together all the small bubbles that appeared when the water broke, but it proved impossible to get sufficient accuracy. (It will be remembered that Dixon and Joly obtained several values by this method, of which the *largest* was only 7·5 atmospheres.)

It was therefore decided to use a weighing method, to determine this volume, analogous to Dumas' method of determining vapour density. An account of a typical experiment follows.

A soda-glass tube of bore 3 mm. and wall thickness 1 mm., and about 15 cm. long, was rinsed with acetone, then with distilled water, and dried. It was then sealed at one end and drawn down to a jet near the other end, annealed in the yellow flame and filled with distilled water by heating and cooling. It was then connected to a filter pump and left at room temperature until no more air bubbles appeared. Those which had appeared were coaxed through the constriction by tapping and by alternately removing the tube from the filter pump and replacing it. With the filter pump still running, the water was driven from the centre of the constriction by a small blow-flame. The heat was then increased and the tube sealed. (The finished tubes were about 10 cm. long, weighed about 5-7 grams and contained about 1 c.c. of water. Some were sealed while open to the atmosphere, but it was found much easier to seal under the pump.) The finished tube was transferred to a beaker of water which was frequently stirred and heated until the bubble in the tube had become fairly small. The rate of heating was then reduced to about $\frac{1}{2}°$ c. per min. Since the time-constant for conductivity of heat through such a tube comes out at about 4 sec., there are unlikely to be any serious temperature differences. It was found possible to determine a "filling temperature" for the tube. If the bath was below the filling temperature, a bubble appeared in the tube almost immediately after it was removed from the bath, and expanded rapidly as the temperature fell. If the bath was above this temperature, the liquid continued to fill the tube for several seconds after removal from the bath, finally failing with a "click" very similar to the cracking of glass. With a little experience it was found that the filling temperature could be determined to within $1°$ c. and in most cases to better than $\frac{1}{2}°$ c. It was found possible to remove the tube from the bath, watch it for a few seconds and decide whether the filling temperature had been reached or not, and then replace it in the bath before a break occurred. The filling temperature was found be to very reproducible. The temperature at which the last visible bubble disappeared could not be determined very accurately, but it appeared to be about $2°$ to $5°$ c. below the filling temperature. The temperature at which the liquid broke was next determined by transferring the tube to a smaller bath (actually a boiling tube), containing water at the same temperature, which was allowed to cool naturally in the laboratory at the rate of $\frac{1}{2}°$ to $1°$ c. per minute, with occasional stirring. The temperature at which the water broke was immediately read and, in most cases, was reproducible to within $\frac{1}{2}°$ c. (To make sure that the rate of cooling was not excessive, one tube was tested by allowing it to cool down in the much larger bath in which it had been heated, the natural cooling rate of which was about $0°\cdot1$ c. per minute. The breaking temperature was twice determined in this way, and the results were indistinguishable from those obtained with the quicker rate of cooling.) The tube was then nicked with a file at the constriction, dried carefully, left in the balance-case to attain a steady state and weighed to $0\cdot1$ milligram with a catenary-chain balance. The tube was then broken open, filled with water by heating and cooling in the bath, and removed from it when the bath had attained the breaking temperature. The tube and piece cut off were dried carefully and again weighed

to 0·1 milligram. (A "blank" experiment showed that, provided the break was clean, there was no measurable loss of weight due to possible splinters of glass.) The difference, usually of the order of 2–3 milligrams, represents, after correction is made for the yielding of the glass, the contraction of the water after failure. The tube was finally weighed empty to determine the total mass of water. At the suggestion of Professor Rideal, four tubes were tested after having been cleaned with hot chromic acid instead of acetone, and in one of these an approximately ½% soap solution was used instead of water. The results were very similar to the others.

A second series of experiments was carried out on identical lines except that a few steel chips or a turning were included in the tube. The steel was washed with acetone, dried by exposure to air, then rinsed with water. The filling tem-

Table 1. Results for plain glass tubes

Filling temp. (° c.)	Breaking temp. (° c.)	Extension (%)		Method of sealing tube	Remarks
		Apparent	True		
82	59	1·35	—	Under atmosphere	
80·5	67·5	0·79	0·34	,, ,,	
75	70·5	0·24	0·11	Under pump	
48	33	0·59	0·14	,, ,,	
41	19	0·61	—	Under atmosphere	
59	36·5	1·03	—	,, ,,	
61	38·5	1·08	0·21	,, ,,	
58	40	0·87	0·12	,, ,,	⎱ Same tube gave two
58	37	0·95	0·19	,, ,,	⎰ sets of values
58	37·5	0·91	0·15	,, ,,	
58	40·5	0·85	0·079	,, ,,	⎱ Different tubes
62	48	0·70	0·20	,, ,,	⎰
54·5	40	0·71	0·20	Under pump	Cleaned with chromic acid
63·5	53·5	0·47	0·12	,, ,,	,, ,,
72	59·5	0·70	0·10	,, ,,	,, ,,
61	50	0·55	0·17	,, ,,	⎧ Ditto with ½% soap ⎩ dissolved in the water

Mean true extension=0·16%, corresponding to 32 atmospheres tension.

perature was not so easy to determine as with the plain water and exhibited a decided tendency to rise, settling down to a value 3° to 5° c. above its initial value after a few cycles. The filling temperature of one tube was determined and again determined 10 days later. It was found to have risen 7° c., but no rust was visible on the steel. This anomaly is no doubt due to the absorption of dissolved gases by the steel, thus increasing the temperature necessary to develop a given pressure in the tube. (It was never observed with the plain water tubes.) The breaking temperature was always very reproducible, but the "clicks" seemed to be noticeably feebler than with plain water. The results are given in tables 1 and 2.

The columns headed "Apparent extension" were obtained from the filling and breaking temperatures, using the data in the International Critical Tables on the density of water at various temperatures. No corrections have been made for the thermal expansion of glass or for its yield under stress. If we had assumed

that the pressure was zero at the filling temperature and had then corrected for the elastic yield, we should obtain tensions comparable with Dixon's (1909).

The columns headed "True extension" were obtained from the weighings described above, in which the percentage contraction of the water after failure was determined. It is necessary to correct these figures for the elastic deformation of the glass, and this was done by the method used by Dixon. The corrected values are those given in the table. These readings do not depend on the filling temperature, and are thus independent of those for apparent extension.

Suppose we have a tension T in the tube, the fractional increase in volume of the water is βT, where β is the coefficient of compressibility of the water. The

Table 2. Results for tubes containing pieces of steel

Filling temp. (° c.)	Breaking temp. (° c.)	Extension (%)		Method of sealing tube	Remarks
		Apparent	True		
72	51	1·09	0·060	Under atmosphere	Steel and brass filings
51·5	40	0·56	—	,, ,,	,, : initial values
58·5	42	0·80	—	,, ,,	,, : values after 10 days
62	51	0·55	0·076	,, ,,	Very small steel chips.
51	43	0·40	0·010	Under pump	Steel turning
72	53	1·01	0·041	Under atmosphere	,, ,,
62·5	48	0·73	0·042	,, ,,	,, ,,
64·5	50	0·74	0·113	,, ,,	,, ,,

Mean true extension $=0\cdot058\%$, corresponding to 12 atmospheres tension.

fractional decrease in the internal volume of the tube is, by Lamé's equation,

$$\left(\frac{\Delta V}{V}\right)_{glass} = \frac{T}{R^2 - r^2}\left(\frac{r^2}{k} + \frac{R^2}{n}\right),$$

where

$R=$ external radius of tube, $\quad k=$ compression modulus of glass,
$r=$ internal radius of tube, $\quad n=$ torsion modulus of glass.

Evidently this quantity must be added to βT to obtain the relative change of volume which is given by the weighings. We have

$$\left(\frac{\Delta V}{V}\right)_{obs.} = \left(\frac{\Delta V}{V}\right)_{water} + \left(\frac{\Delta V}{V}\right)_{glass} = \beta T + \frac{T}{R^2 - r^2}\left(\frac{r^2}{k} + \frac{R^2}{n}\right). \quad \ldots\ldots (1)$$

$\Delta V_{obs.}$ and $V_{obs.}$ are determined directly from the weighings, independently of the temperatures.

We have

$\beta = 49\cdot5 \times 10^{-6}$ per atmosphere. (Kaye and Laby.)
$k = 4 \times 10^5$ atmospheres. (Dixon.)
$n = 3 \times 10^5$ atmospheres. (Dixon.)
$R = 0\cdot25$ cm. $r = 0\cdot15$ cm.

Inserting these values, we obtain the results

$$\Delta V/V_{obs.} = 56 \times 10^{-6}\, T \text{ (T being measured in atmospheres).}$$

$$\left(\frac{\Delta V}{V}\right)_{water} = 0.87 \left(\frac{\Delta V}{V}\right)_{obs.} = 49.5 \times 10^{-6}\, T.$$

$\left(\dfrac{\Delta V}{V}\right)_{water}$ is tabulated.

Strictly speaking we ought to allow for the variation of β with temperature. Inspection of Bridgman's data in the International Critical Tables indicates that this variation is fairly slow (actually a decrease with rising temperature), but exact figures cannot be obtained as Bridgman uses a large pressure-interval :—

It will be seen, therefore, that we obtain the following values for tension in water, using the corrected figures for true extension :—

In the presence of glass. Mean 32 atmospheres. Maximum 68 atmospheres. Minimum 16 atmospheres.

In the presence of steel. Mean 12 atmospheres. Maximum 23 atmospheres. Minimum 2 atmospheres.

It was thought advisable to obtain an independent check on the validity or otherwise of Dixon's assumption of zero pressure at the filling temperature. One method of doing this was to examine the tube between crossed polaroids, which enabled the changes of stressing of the tube to be followed qualitatively, and a rough determination to be made of the temperature at which the pressure in the tube was zero. A second method, suggested by Sir Geoffrey Taylor, was to enclose the tube in an outer jacket connected to a capillary, which enabled the change in volume of the Berthelot tube when it was cut open, or when it was allowed to cool until the liquid broke, to be determined directly. While it can be stated that the results confirm those of the present investigation, it has been thought advisable to publish the details separately.

§ 6. DESCRIPTION OF MODE OF FAILURE OF WATER

As already stated, the failure occurred with a click. The water remained "crystal clear", even when examined by very strong light, right up to the moment of failure, there being no sign of opalescence such as occurs near the critical point. The first appearance at failure was a cloud of barely visible bubbles, which cloud faded away in about half a second, leaving a small number of bubbles about $\frac{1}{2}$ mm. in diameter. For the tubes with steel in them the cloud invariably occurred near the steel. For the tubes containing plain water it was difficult to say whether the failure occurred in the water or at the glass surface. The tubes were kept standing upright while cooling and they sometimes jumped up a distance of one or two centimetres at failure. The energy associated with this is of the order of 10^4 ergs, a minute fraction of the energy stored up in the liquid under tension. The figure becomes of a more reasonable order of magnitude when one reflects that not all of this energy would go into sound waves and that only the energy directly incident on the bottom of the tube would be available for lifting it, the remainder travelling outwards into the water in the bath.

§7. EXAMINATION OF THE REYNOLDS CENTRIFUGAL METHOD OF MEASURING TENSION

After a number of trials, an apparatus was developed consisting essentially of a disc of aluminium rotated by an electric motor. A U-tube was clamped with its long arm along the diameter of the disc and its short arm lying flat on the disc. The long arm was sealed, and the short arm was open to the atmosphere. The speed of rotation was measured in three distinct ways:

(*a*) By tachometer.

(*b*) The motor was really a motor-generator and had two distinct commutators and armature windings. By connecting a voltmeter across one armature and the supply across the other one could measure the voltage developed and obtain the speed once the voltmeter had been calibrated against the tachometer. This calibration was not sensitive to change in voltage of the supply mains.

(*c*) Late in the series of experiments a " Strobotac " (Neon tube with variable frequency of flashing) was kindly loaned by E. in C. Department, Admiralty, which enabled the speed of rotation to be measured very accurately. It was hoped that it would also be possible to observe the cavitation process, but it was too rapid to follow visually.

The disc was 9″ in radius which, after allowing for the clamping arrangements and the length of the column of water in the short arm, gave an effective radius from centre to meniscus of about 8″. The disc could be rotated at speeds of up to 2000 r.p.m., which gave a maximum tension of about 8 atmospheres if it may be assumed that the water moves like a rigid body so that we may apply the ordinary law of centrifugal force. (We shall discuss this assumption later.) Evidently the pressure at the centre is given by

$$p = p_0 - \tfrac{1}{2}\rho\omega^2 r^2,$$

where p_0 is the pressure of the atmosphere, r the effective distance between the centre and the meniscus and ω the angular velocity. Since all these quantities can be measured, the tension at the centre just before break can be calculated. (The velocity of rotation was increased slowly by means of a variable resistance in order to avoid any contribution to the tension from terms involving $d\omega/dt$. It is easily shown that the Bernoulli variation of pressure due to some air being carried round by the disc cannot be large enough to matter.

Attempts were first made to use air-free water, but we found, by a sensitive test to be described later, that, even if water is boiled for many hours in the atmosphere or is left under a vacuum with constant shaking, there are still dissolved gases left. Since the readings for a single sample of water showed wide variations of critical tension it would be difficult to compare the readings on two different days if we used water with uncertain amounts of dissolved gas in it. It was therefore decided to use tap water without treatment. It was found to be saturated, or nearly saturated, with air, as would be the water in many cases in which cavitation occurs in practice. The experiments were carried out with four different bores of tube (1 mm., 2 mm., 3 mm. and 5 mm.). Late in the series it was realized that the distance between the axis of rotation and the sealed end of the tube might also be of significance, so the overall length of the tube was recorded.

§ 8. RESULTS

A summary of the results is given in table 3.

Table 3

Bore of tube (mm.)	No. of experiments	Mean tension (atm.)	Highest reading (atm.)	Lowest reading (atm.)
5	16	0·88	2·2	0·1
3	25	1·23	5·3	Zero
2	24	1·00	5·0	Zero
1	27	2·34	5·65	Zero

For those experiments in which it was recorded, the overall length of tube was plotted against critical tension. It was obvious that the correlation between tension and length is very slight, even if it exists at all. There is, however, clear evidence of an increase of critical tension as the bore of the tube is decreased from 5 mm. to 1 mm., though the results for 3 mm. and 2 mm. show a slight reversal. As we shall see, there are theoretical reasons for expecting such an increase in observed tension with smaller diameter. It will be seen that the highest values are of the same order of magnitude as that obtained by Reynolds (4·8 atmospheres), though we do not know what bore of tube he used.

§ 9. THEORETICAL ANALYSIS OF THE CENTRIFUGAL METHOD

On the assumption that the motion of the liquid in the tube is entirely irrotational, we can discuss it qualitatively by using the well-known equations for motion in an ellipsoidal envelope (Lamb's *Hydrodynamics*, 5th edn. p. 147). This solution implies considerable slipping of the liquid along the boundary near the short axis of the ellipsoid, but it also implies that, in any real liquid, there must be a boundary layer growing with time, and that in time all the liquid in the tube will be rotating with the same angular velocity as the boundary. (We are very grateful to Sir Geoffrey Taylor for pointing this out to us.) The time interval associated with the change-over is easily computed from the equations of diffusion of vorticity, which are of the same form as the equation of heat-conduction. It is of the order of $\eta a^2/\rho$, where η is the viscosity, ρ the density, and a is the radius of the tube. For a tube of 5 mm. bore this is only about 6 seconds, which is short compared with the time taken by an experiment, so the stirring-up of the liquid due to its inertia is not likely to be important.

If however, the angular velocity is not constant, but increases with time, we still get a stirring effect even if the liquid as a whole does not lag behind the rotation of the tube. If we reduce the tube to rest by a transformation of axes, we have to introduce the following field of force to allow for the effect of the rotation

$$F = m\left(\omega\wedge[\omega\wedge\mathbf{r}] + 2\omega\wedge\frac{d\mathbf{r}}{dt} + \frac{d\omega}{dt}\wedge\mathbf{r}\right), \qquad \ldots\ldots(2)$$

where m is the mass of a particular particle of fluid, $\mathbf{r}$ its radius vector, and ω the vector representing angular velocity, which we may suppose to be directed along

the z axis. The first term in this expression is the ordinary centrifugal force which produces a pressure-gradient along the tube. The second term is the Coriolis force, which can contribute nothing to the pressure-gradient along the tube unless there is relative motion of the tube and the liquid as a whole.   The third term is proportional to the angular acceleration and gives a longitudinal pressure-gradient antisymmetrical about a plane through the z axis and the axis of the tube. The corresponding two-dimensional problem can be solved easily. Suppose we have the liquid moving between two planes at a distance d apart. Let the x axis lie midway between the planes and let the y axis be perpendicular to them. Then we have

$$\frac{\eta}{\rho}\frac{\partial^2 V_x}{\partial y^2} = \frac{1}{\rho}\frac{\partial p}{\partial x} = \frac{d\omega}{dt}\,y, \qquad \ldots\ldots(3)$$

and the solution of this equation for which the tangential velocity vanishes at $y = \pm d/2$ is

$$V_x = \frac{\rho}{6\eta}\left(y^3 - y\frac{d^2}{4}\right)\frac{d\omega}{dt} \qquad \ldots\ldots(4)$$

The maximum velocity occurs at $y = \dfrac{d}{\sqrt{12}}$ and is given by $V_x = \dfrac{d^3\rho}{36\eta\sqrt{12}}\dfrac{d\omega}{dt}$.

This implies that if the tube is slowly accelerated from an angular velocity of zero to ω, a particle of liquid may be shifted a distance of the order of magnitude $\dfrac{\rho\omega d^3}{36\eta\sqrt{12}}$, if d is the diameter of the tube. For our widest tubes d was 5 mm., which makes this distance 13 cm. for a final rate of rotation of 20 revolutions per second, and this is comparable with the total length of the tube. It is therefore quite likely that minute air-bubbles or other nuclei might be carried to the axis of rotation and thus help the liquid to break. (This theory also we owe to Sir Geoffrey Taylor.)

Our observed tensions are in fact larger for the smallest tubes, in agreement both with this and the irrotational theories, but the irrotational theory also predicts a larger observed tension when the axis of rotation is near the sealed end of the tube than when it is near the centre, because of the greater slipping that would occur in the latter case, whereas there is no particular reason for expecting such correlation on Sir Geoffrey Taylor's theory.   In fact we looked for correlation between the observed tensions and the distances of the axis of rotation from the sealed end, but we could find no evidence of any.

There is a second mechanism that would tend to bring any minute gas bubbles towards the axis of rotation—the pressure gradient in the tube. As we have seen, the distribution of pressure is given with sufficient accuracy by

$$p = p_0 - \tfrac{1}{2}\rho\omega^2 x^2.$$

The force on a bubble of radius r due to the pressure gradient is given by

$$\frac{4}{3}\pi\rho r^3 \frac{dp}{dx}.$$

The force on it due to viscosity is given by

$$6\pi\eta r \frac{dx}{dt}.$$

(Stokes' law is here only a rough approximation owing to the presence of the walls.) The force on the bubble due to acceleration of the surrounding liquid is

$$\frac{2}{3}\pi\rho r^3\frac{d^2x}{dt^2},$$

since the virtual mass is half that of the displaced liquid.

The equation of motion is, therefore,

$$\frac{d^2x}{dt^2} + \frac{9\eta}{2\rho r^2}\frac{dx}{dt} + 2\omega^2x = 0. \qquad \ldots\ldots(7)$$

Consider a small bubble of radius such that its surface tension just balances the centrifugal tension on the axis of rotation. Assume the values for the constants already taken, i.e. $a = 20$ cm., $\omega = 40\pi$. The tension at the centre is the 2·15 atmospheres, and the corresponding value of r is $6\cdot8 \times 10^{-5}$ cm., taking the surface tension as 74 dynes/cm. Taking $\eta = 0\cdot01$ c.g.s. units and inserting numerical values, equation (7) becomes

$$\frac{d^2x}{dt^2} + 19\cdot4 \times 10^6\frac{dx}{dt} + 3\cdot2 \times 10^4x = 0,$$

whence it is clear that the motion of such a bubble would be over-damped and would have a time-constant of the order of 600 seconds.

The conclusion that one reaches from these investigations is that the centrifugal method is theoretically unfavourable for the production of large tensions even if there is no vibration in the apparatus. Any vibration would set up acceleration and reduce the observed tension, because any small bubbles previously existing in the liquid would tend to be brought near the axis of rotation by the lag of the water behind the tube and by the pressure gradient of the centrifugal force, and this process might well take a time comparable with that needed to do the experiment, especially for a very narrow tube. Harvey found that very vigorous centrifuging (acceleration $4000\,g$) is necessary to remove gas-nuclei from water in a reasonable time. (The accelerations used by us in the above experiments were only about $500\,g$.)

§ 10. A STUDY OF CAVITATION IN A CONSTRICTED TUBE

A simple apparatus was constructed consisting of a pressure-flask exhausted by a filter-pump, to which were connected a mercury manometer, a stop-cock open to the atmosphere, enabling the pressure to be held at any desired value, and a tube with a constriction connected to the pressure flask by a short piece of rubber tubing. The other end of the tube was immersed in a beaker of the water to be tested. The pressure in the flask was allowed to fall until cavitation just occurred at the constriction. The pressure-difference was then read on the manometer and held constant while the rate of flow of water was measured. A "blank" experiment was then performed with a tube without a constriction, the pressure being adjusted until the rate of flow was the same. At the conclusion of a series of experiments, the tube was cut open at the constriction and its diameter measured by a travelling microscope. The tension can now be calculated if it may

be assumed that the flow in the constriction is laminar. If it is turbulent, the actual tension will be greater. An analysis of a typical set of experiments follows:

Bore of tube 0·3 cm.

Mean diameter at constriction 0·078 cm. Area $= 4·8 \times 10^{-3}$ cm.²

Tap water.

Cavitates at 50·5 cm. below atmosphere and 4·2 c.c./sec. flow $= 875$ cm./sec.

Calculated tension, 0·05 atmosphere.

Tap water passed once through apparatus.

Cavitates at 58·0 cm. below atmosphere and 4·3 c.c./sec. flow $= 895$ cm./sec.

Calculated tension, 0·22 atmosphere.

Tap water passed 3 times through apparatus, then left 3 hours under vacuum.

Cavitates at 64·5 cm. below atmosphere and 4·4 c.c./sec. $= 920$ cm./sec.

Calculated tension, 0·27 atmosphere.

The Reynolds number for such rates of flow is of the order of 4000, so the flow may be turbulent, in which case these values are lower limits only. A region of turbulence would obviously be a very unfavourable one for the survival of a metastable state of the liquid, so these values cannot fairly be compared with those obtained under static conditions.

The apparatus was designed with two other objects in mind:

First, to detect the presence of dissolved gases in water. This can easily be done by sucking the water through the constriction until it cavitates, then suddenly stopping the flow by pinching the rubber tube. The pressure rises to atmospheric, the cavitation disappears except for any dissolved gases that may have been released by the cavitation, which remain in the form of bubbles. Tests on water which had been boiled for many hours still showed traces of gases, and the same was true of water left in a vacuum for several hours or water that had passed many times through the apparatus, so it appears to be very difficult to remove nuclei completely from water.

The second object was to try and get a qualitative picture of what probably occurs behind a " breaking-front " in water, e.g. when an explosion pulse is reflected at a free surface, causing cavitation. The direction of flow of water is in general *away* from such a front. For this study a tube about 3 feet long, with a constriction near the far end, was used. For water nearly saturated with air, isolated bubbles appeared a long way beyond the constriction long before they were in evidence near it. When cavitation at the constriction was definitely established, the sizes of bubble just beyond it were distributed almost at random, but the large bubbles moved faster and overtook the smaller ones, so that the distribution showed large bubbles far from the constriction, and small ones near it.

§ 11. DISCUSSION

We think that our examination of Reynolds' centrifugal method gives an acceptable explanation of the low values commonly obtained by it. Our examination of the Berthelot method seems to account for the high readings obtained by Dixon (1909), but it remains to be explained why it needs a high pressure (which from our experiments appears to be of the order of 50–100 atmospheres)

to persuade the water to stick to the glass. It was suggested to us by Professor Rideal (private communication) that this might be due to the trapped gas being forced into fissures in the glass. Harvey (1944) has pointed out that if the "advancing" and "receding" contact angles differ, there will, for any given mass of gas in a fissure, be a finite range of pressures for which the size of the gas bubble in the fissure will remain practically constant. However, it does not follow that one would get larger tensions by increasing the intial pressure, because one might alter the shape of the fissures. Vincent's work (Vincent and Simmonds, 1943, p. 378) clearly shows that there is a limit to the excess pressure one can usefully apply, and with our tubes which are near their bursting pressures the limit is probably much lower. One of our tubes was by mistake heated 5° c. above the filling temperature, but the breaking temperature was unaltered. A repetition of the experiment gave a similar result.

If we discount the high readings obtained by Dixon and the low values obtained by the Reynolds centrifugal method we have the following set of values for water in the presence of glass:

16–68 atmospheres (mean 32 atmospheres).   (Modified Berthelot.)
34 atmospheres.   (Meyer.)

For water in the presence of steel we have:

2–23 atmospheres (mean 12 atmospheres).   (Modified Berthelot.)
Up to 60 atmospheres.   (Budgett.)

The agreement for the second set of results is still not very good.

There remain the very low values obtained by Vincent, with his tonometer (7·8 atmospheres) and his metal bellows (2–3 atmospheres). Though none of these are for water, they are for liquids which seem to behave like water in that by other methods they have given results comparable with water. The low values with the bellows may well be due to the fact that pressure was not previously applied. Our results show also that low values are quite possible in the presence of steel even *with* the prior application of pressure. The tonometer suffers from one of the defects of the Reynolds method in that there is a gradient of tension which would tend to drive any nuclei towards the regions of greatest tension, and this may account for the low values observed with it.

§ 12. CONCLUSION

Ordinary water, even if it contains dissolved air, can stand static tensions of the order of 40 atmospheres. Under dynamic conditions the value is probably less, and drops to an almost negligible value if the flow becomes turbulent. The work with the Reynolds apparatus establishes that even water nearly saturated with air can stand tensions of up to 6 atmospheres. A tension of 40 atmospheres is ample to pull sap up the tallest trees, and is not inconsistent with the data on the projection of spores.

ACKNOWLEDGMENTS

REFERENCES

BUDGETT, 1912. "The adhesion of flat surfaces". *Proc. Roy. Soc.*, A, **86**.
DEAN, 1944. "The formation of bubbles." *J. Appl. Phys.*, **15**, 446.
DIXON, 1909. *Sci. Proc. Roy. Soc. Dublin*, **12**.
DIXON and JOLY, 1895. "The ascent of sap." *Phil. Trans.*, **186**, 563.
FÜRTH, 1941. "On the theory of the liquid state." *Proc. Camb. Phil. Soc.*, **37**, 276.
HARVEY *et al.*, 1944. "Bubble formation in animals." *J. Cellular Comparative Physiol.*, **1**, 23.
JOFFÉ, 1928. "Discussion on cohesion." *Trans. Faraday Soc.*, **24**.
KENRICK, GILBERT and WISMER, 1924. "The superheating of liquids." *J. Phys. Chem.*, **28**, 1297.
KENRICK, WISMER and WYATT, 1924. "Supersaturation of gases in liquids." *J. Phys. Chem.*, **28**, 1308.
KING, 1944. "The spore discharge mechanism of common ferns." *Proc. Nat. Acad. Sci. Wash.*, **30**, 155.
KORNFELD and SUVAROV, 1944. "The destructive action of cavitation." *J. Appl. Phys.*, **15**.
MEYER, 1911. *Abh. Deutschen Bunsen Ges.*, **6**.
REYNOLDS, 1878. *Mem. Manchr. Lit. Phil. Soc.*, **17**, 159.
VINCENT, 1941. "Measurement of tension in liquids by means of a metal bellows." *Proc. Phys. Soc.*, **53**, 26 ; 1943 a. "A new method of measuring tension in liquids." *Ibid.*, **55**, 41.
VINCENT and SIMMONDS, 1943 b. "Examination of the Berthelot method of measuring tension in liquids." *Proc. Phys. Soc.*, **55**, 376.
WORTHINGTON, 1892. *Nature, Lond.*, **42**, 261.

Note added in proof: I no longer consider that the theoretical breaking strength of a liquid is equal to the so-called "intrinsic pressure." It may, in fact, be very much less. I hope to discuss this point in a later paper.

H. N. V. T.

THE BEHAVIOUR OF WATER UNDER HYDROSTATIC TENSION : II

BY H. N. V. TEMPERLEY,
King's College, Cambridge

ABSTRACT. In continuation of the research described in Part I (Temperley and Chambers, 1946), a further examination of the Berthelot method of producing tensions in liquids has been made. The critical tension of water has been measured by a method which does not assume that the extensibility and compressibility of water are equal, and other evidence in confirmation of this assumption has been obtained. It is concluded that high pressures are necessary in the Berthelot tube to force the final gas bubble to dissolve in a reasonable time, and that this fact is probably due to the low rates of diffusion of gases through liquids. It is also concluded that there is a large discrepancy between the theoretical and the observed strength of water.

§ 1. INTRODUCTION

IN Part I of this paper (Temperley and Chambers, 1946) some experiments on the behaviour of water under tension were described. Two methods of producing tension in liquids were examined in detail, the Reynolds centrifugal method and the method due to Berthelot, which consists of sealing water or other

liquid in a glass tube together with a trace of air, heating the whole until the liquid expands sufficiently to fill the tube completely, then cooling until the liquid fails by hydrostatic tension and no longer fills the tube completely. This method has been used by a number of observers besides Berthelot, for example Dixon and Joly (1895), Dixon (1909) and Vincent (1943). In Part I the available evidence was reviewed and some experiments by a modification of the Berthelot method were described, and it was concluded that the very high values of critical tension, of the order of 200 atmospheres, sometimes obtained by this method were probably false, and that the true value was probably in the neighbourhood of 30 to 50 atmospheres. It was suggested that the discrepancy was probably due to high pressures being generated in the tubes before they filled with liquid. If this conclusion is correct, it means that the theoretical strength of water is too high by a factor of the order of 100. A discrepancy of similar order has long been known to exist for solids, but the conventional explanations of this in terms of surface-cracks and stress-concentrations seem quite inadmissible for liquids. In view of the possible theoretical importance of the conclusions of Part I, it was thought desirable to obtain independent confirmation of them by at least one new method. This paper describes a further examination of the Berthelot method and some new modifications of it.

§ 2. PRELIMINARY EXPERIMENTS

A number of tubes were made up by the method described in Part I with " filling temperatures" ranging from 30–80° c. (It will be remembered that the "filling temperature", as defined in Part I, was the lowest temperature at which a tube would fill completely and thus enable the water to be subjected to tension on cooling.) These tubes were heated to the filling temperature in a beaker of water which was then allowed to cool slowly. The state of stress in the glass could be followed qualitatively by viewing the tube between crossed pieces of Polaroid, the glass becoming doubly refracting under stress, and it was invariably found that the glass was under stress at the filling temperature, that the stress decreased to a small or zero value as the temperature fell, only to increase again as the breaking temperature was approached, finally vanishing abruptly when the water inside the tube broke under tension. Unfortunately, it was not possible to locate with any accuracy the temperature at which the stress in the tube was zero, there being an uncertainty in either direction of the order of 2 or 3° c., corresponding to about 10 to 20 atmospheres. This photo-elastic method in its present form could, therefore, only give very rough measurements of tension, but it does provide qualitative confirmation of the work described by Temperley and Chambers in Part I, by proving that the pressure in the tube is *not* small at the filling temperature, and that the temperature at which the pressure does vanish is nearer the breaking temperature than the filling temperature.

The actual process of disappearance of the bubble of gas as a tube was heated slowly to the filling temperature was watched by means of a low-power microscope (magnification about 30). It would have been of little use to employ a higher power, because the image of the bubble became blurred as the filling temperature was approached, presumably owing to the stresses destroying the optical uniformity of the glass. It was, however, possible to identify definitely the instant at

which the last bubble disappeared, and it was found that, if the tube was immediately removed from the bath, the water subsequently broke under tension, so that the temperature of disappearance of the bubble was identical with the filling temperature. (In the work described in Part I the disappearance of the bubble was judged visually, but without a microscope, and the two temperatures were thought to be separated by a few degrees.) It is hard to see why a minute gas bubble should be able to persist for long periods in the presence of liquid at high pressure, but this appears to be due to the very low numerical values of diffusivities. We shall return to this point in more detail later on.

It was thought possible that something analogous to the cloud-chamber phenomenon might occur with water in its metastable state under tension, in that an energetic charged particle might act as a nucleus and initiate the failure of the liquid. Various tubes have been cooled in the presence of a beta-particle source, but the results have so far been completely negative.

Attempts were also made to modify the Berthelot experiment for water by lowering tubes gradually into brine kept at about $-10°$ c. by a freezing mixture, instead of by heating them. The expansion of the water as it froze set up pressure and compressed the gas bubble remaining in the tube, exactly as in the ordinary version of the experiment, but the freezing had to be done much more cautiously than the heating in order to get the tube to fill without bursting. It was found that water adheres to ice less strongly than it does to glass, in that failure (when the tube was removed from the brine and allowed to warm up naturally in the laboratory) invariably occurred by the formation of a few fairly large bubbles at an ice-water surface and without the sharp click that is so characteristic of the ordinary Berthelot experiment. One might interpret this result as meaning that water cannot exert tension on ice in equilibrium with it. One might expect water to adhere to ice better than to other solids, but the result becomes reasonable when one reflects that, at pressures less than that of the triple point, the ice would sublime into vapour directly, so that the bubbles formed may have come directly from the ice. The experiment, then, may simply mean that ice cannot assume a metastable state under tension in the way that water can. (It is well known that, while water can be supercooled to a considerable degree, ice cannot be super-heated.)

§3. QUANTITATIVE MEASUREMENTS

A serious defect of the work described in Part I was that taking a measurement necessarily involved the destruction of the tube when it was cut open in order to ascertain (by a weighing method) the total volume of cavity formed by the liquid when it broke. The reproducibility or otherwise of the critical tension associated with a given tube could then only be inferred indirectly from the reproducibility or otherwise of the breaking temperature. It will be remembered that in the work described in Part I the breaking temperature was found to be very reproducible except for one tube.

The method used was suggested by Sir Geoffrey Taylor, and it is similar in principle to that used by Worthington (1892). The Berthelot tube was surrounded by a water-jacket which ended in a fine capillary which was open to the atmosphere and was furnished with a scale to enable the movement of the meniscus to be followed. An experiment was carried out as follows: The whole was heated

just above the filling temperature for the Berthelot tube and was then cooled to a temperature a few degrees above the breaking temperature, the open end of the capillary dipping into a reservoir so that the whole jacket was full of water. The reservoir was then removed and the cooling continued. The meniscus gradually fell in the capillary, and its position was read every half-minute and a cooling curve plotted. At the breaking temperature, the stress on the glass of the Berthelot tube was suddenly released as the tension vanished, and the tube consequently increased in volume, as was shown by a sudden rise in the level of the meniscus, and a consequent discontinuity in the cooling curve. The size of this discontinuity, together with a knowledge of the cross-section of the capillary and the dimensions of the Berthelot tube, enabled the tension inside the Berthelot tube to be estimated.

A slight complication of this method is that the result would be vitiated if there were a large calorimetric effect when the liquid broke, because the heat released would cause the water in the jacket to expand, and this would be superposed

Table 1.   Summary of results with first tube

Group	No. of expts.	Mean vol. change (10⁻⁴ c.c.)	Estimated tension (atmospheres)			Liquid in jacket
			Mean	Max.	Min.	
A	8	3·9	18·5	27	13·5	Water
B	11	7·8	37	44·5	25	Water
C	5	9·9	46·5	53·5	31·5	Meth. spirit

on the effect due to the change of volume of the Berthelot tube. Although consideration of the various heat capacities involved showed that the error would not be serious even if all the energy stored up as tension in the water appeared as heat, it was considered desirable to make an estimate of the order of magnitude of the calorimetric effect. This was done by using methylated spirit in the jacket instead of water. Methylated spirit has a smaller heat capacity per unit volume, but a much larger coefficient of expansion than water, thus increasing the contribution of any calorimetric effect to the observed volume change. To make this contribution as large as possible, the tube and jacket were allowed to cool in the laboratory, the outer bath being removed altogether. The rate of fall of the meniscus in the capillary was kept small by using a Berthelot tube whose breaking temperature was only just above room temperature. Such tubes are quite easily made by the technique of sealing-off under a filter-pump described in Part I. A summary of the results obtained with this tube is given in table 1.

Details of apparatus

Berthelot tube: Effective length 10 cm. Internal radius 0·19 cm. External radius 0·29 cm. Filling temperature 47° c. Breaking temperature 18°·5 c. Temperature of zero pressure estimated photoelastically to be $26·5 \pm 2°$ c. Water equivalent 0·55 gm. Internal volume 1·17 c.c. External volume 2·55 c.c.

Jacket: length of wide portion 11·5 cm. Length of capillary 65 cm. Weight 12 gm. Water equivalent 1·9 gm. Volume (available for liquid in jacket with

Berthelot tube in place) 1·77 c.c. Cross-section of portion of capillary used varied between 4·1 and 5·9 × 10⁻⁴ cm².

In group A the volume change was estimated from direct observation of the meniscus; in groups B and C it was deduced from cooling curves. Owing to the small cross section of the capillary, the movement of the meniscus was sometimes sluggish, and it continued to rise for several seconds after the water in the Berthelot tube broke. It is thus not surprising that the results in group A are low. The systematic difference between groups B and C does suggest that there was a calorimetric effect, a hypothesis consistent with the shape of some of the cooling curves. (Group C was brought prematurely to an end by the bursting of the tube.)

The tensions were estimated by means of Lamé's formula (see Love, *Elasticity*).

$$\frac{\Delta V}{V} = \frac{T}{R^2 - r^2}\left(\frac{r^2}{k} + \frac{R^2}{n}\right),$$

where k is the bulk modulus of glass (4×10^5 atm.) and n is the rigidity modulus (3×10^5 atm.), these values being the ones used by Dixon (1909). We can estimate the order of magnitude of the calorimetric effect in the following way: suppose that x calories are liberated per c.c. of water in the Berthelot tube. If we assume that this heat is shared between the water in the tube, the tube itself and the liquid in the jacket, the rise in temperature is found to be $0·335\,x\,°$ c. with water in the jacket and $0·456\,x\,°$ c. with methylated spirit in the jacket. If, however, we suppose that the glass jacket also shares the heat, the figures become respectively $0·213\,x$ and $0·256\,x\,°$ c. We assume the following constants for water and methylated spirit. Water: coefficient of expansion in neighbourhood of 20° c. 20×10^{-5}. Methylated spirit: coefficient of expansion 12×10^{-4}; density 0·8; specific heat 0·6. On this basis we find, by a simple calculation, that to account for the observed difference between the means of groups B and C we must have

       $x = 0·24$ calories/c.c. if heat is not shared with glass jacket,

       $x = 0·43$ calories/c.c. if heat is shared with glass jacket.

The potential energy associated with a tension of 37 atmospheres is 0·88 calories/ c.c., so that we may conclude that between 25 and 50% of this eventually appears as heat. Presumably most of the energy appears initially as sound, but is degraded by repeated reflection at the walls of the tubes. Taking a figure of 40%, we see that the values of tension in group B have to be reduced slightly to allow for the calorimetric effect. The corrected values are: Maximum 42 atmospheres; minimum 24 atmospheres; mean 35 atmospheres. These should be compared with the value of 32 ± 10 atmospheres deduced from the photoelastic measurement on the same tube (making the conventional assumption that compressibility and extensibility are equal) and with the values of 16 to 68 atmospheres (mean 34 atmospheres) recorded in Part I, which also involve this assumption, but were obtained from twelve different tubes. The method just described does not involve any such assumption, so that it seems that we have obtained an indirect check on this assumption, but there is still the uncertainty arising from the fact that the check is made by comparing the results from different tubes.

Quite by chance, one of the tubes made during this investigation was found to have a very variable breaking temperature, which provided an opportunity

to decide between this assumption and the alternative hypothesis that microscopic bubbles are formed as soon as tension sets in, thus increasing the apparent extensibility. The tube was fitted with a water jacket, and the volume changes at "break" were determined exactly as before, except that this time the tube and jacket were enclosed in a beaker of water. The whole was allowed to cool slowly, and the temperature of the bath at break was recorded. A blank experiment

Table 2.   Results for tube with variable breaking temperature

Breaking temperature	Number of experiments	$10^4 \times$ observed volume change (c.c.)	$10^4 \times$ calculated volume change (c.c.)	Estimated tension (atmospheres)
38	1	4·0	5·5	19
40	2	4·15	4·3	19·5
40·5	1	4·1	4·05	19·5
41·5	2	4·0	3·5	19
43	4	2·4	2·6	11·5
44·5	1	2·6	1·7	12·3

showed that the time-lag between jacket and bath was quite small, and a simple calculation shows that the time-constant for heat conduction through the walls of the Berthelot tube is also of the order of a few seconds only. Although the results are obviously not very accurate, they do seem to point very definitely to the compressibility and extensibility being equal, and they are accordingly recorded in table 2.   Worthington (1892) reached the same conclusion for ethyl alcohol.

Details of second tube

Effective length 10 cm.   Internal radius 0·19 cm.   External radius 0·29 cm. Filling temperature 58° c.   Breaking temperatures between 44°·5 and 38° c Temperature of zero pressure found photoelastically to be between 45° and 50° c.

The figures headed "observed volume change" refer to the change of the external volume of the Berthelot tube, which was determined by the method already described. The figures headed "calculated volume change" were calculated on the basis that the temperature of zero pressure was 47°·5 c. and that the compressibility and coefficient of expansion of water are the same under tension as at zero pressure. The change in volume was calculated by Lamé's formula as before. The agreement between calculated and observed values is only rough, but does seem definitely to rule out the possibility of any big change in the compressibility of water when tension is applied. As the method is obviously not capable of giving accurate results, an extended series of observations was not considered worth while.

§4. DISCUSSION

We seem to have established that it is necessary to develop pressures of the order of 50 to 100 atmospheres inside the Berthelot tubes in order to cause the final bubble of gas to disappear in a reasonable time, and that this fact has vitiated

some of the earlier measurements. We have now to find some explanation of this fact. We consider the diffusion of gas from the neighbourhood of the bubble. For simplicity, we assume spherical symmetry, and neglect any effect of the glass walls. The equation governing the concentration changes is

$$\frac{\partial c}{\partial t} = n\left(\frac{\partial^2 c}{\partial r^2} + \frac{2}{r}\frac{\partial c}{\partial r}\right). \qquad \ldots\ldots(1)$$

The *International Critical Tables* give n as of the order of 10^{-5} for all common gases (measured in c.g.s. units). We take $c=1$ to mean saturation, and assume boundary conditions $c=1$, $\frac{\partial c}{\partial r}=0$ at $r=a$, the surface of the bubble, where a is a slowly varying function of time. To obtain an estimate of the lifetime of a bubble of radius a, we may, to a first approximation, assume a to be constant, and calculate the rates of decay with time of those solutions of equation (1) which satisfy the boundary conditions initially. This will give us a measure of the rate at which gas diffuses away from the neighbourhood of the bubble. The relevant. solutions of equation (1) are of two types,

$$S_k = \frac{\sin kr}{r}\exp{(-nk^2 t)}, \quad C_l = \frac{\cos lr}{r}\exp{(-nl^2 t)}. \quad \ldots\ldots(2)$$

In order that the derivatives of these solutions should vanish at $r=a$, we must have $ka=\tan ka$ and $la=-\cot la$. The solution which decays with time the least rapidly is given by $la\sim 2$, so that our estimate of the time-constant for the diffusion of gas away from a bubble of radius a is $\frac{10^5 a}{4}$ seconds if a is measured in centimetres. To reduce this time-constant to a reasonable value, say 10 seconds, it follows that we must compress the gas-bubble to a radius of a few tenths of a millimetre, so that the inferred pressures of the order of 50 atmospheres needed for the tubes to fill in a reasonable time are quite reasonable.

A microscopic examination of the process of filling of tubes did show that, when more than one bubble was initially present, the small ones disappeared long before the large ones, and also that, if the bubble was confined in the narrow neck of the tube, diffusion thus being hindered, the tube was much more difficult to fill than if the bubble was shaken into the wider portion. Investigation at temperatures slightly below the filling temperature indicated that the bubbles did still diminish in size, although very slowly, so that reduction of the filling temperatures might be achieved at the expense of prolonging the time of the experiments. The above numerical estimate of the time-constant of the bubbles takes no account of any possible effect of pressure on diffusivity, but increase of pressure is not likely to *increase* the diffusivity. The author has been unable to find any numerical data on this point.

§5. CONCLUSIONS

We therefore seem to be forced to the conclusion that the strength of water in the presence of glass is of the order of 30 to 50 atmospheres, and that the higher values reported by Dixon are due to the high pressures required to make the tubes fill in a reasonable time, which may, in turn, be traced to the low numerical

values of diffusivities in liquids. The discrepancy between the theoretical and the observed strength of water seems to be confirmed, and to merit further theoretical investigation.

ACKNOWLEDGMENTS

REFERENCES

DIXON and JOLY, 1895. *Phil. Trans.*, **186**, 563.
DIXON, 1909. *Sci. Proc. Roy. Soc. Dublin*, **12**.
TEMPERLEY and CHAMBERS, 1946. *Proc. Phys. Soc.*, **58**, 420.
VINCENT, 1943. *Proc. Phys. Soc.*, **55**, 376.
WORTHINGTON, 1892. *Nature, Lond.*, **42**, 261; *Phil. Trans.*, **183**, 355.

THE VARIATION OF RESOLUTION WITH VOLTAGE IN THE MAGNETIC ELECTRON MICROSCOPE*

By V. E. COSSLETT,

Electrical Laboratory, Oxford

ABSTRACT. The spherical and chromatic aberrations, the diffraction error and the total aberration of a magnetic electron lens are discussed, as regards variation with voltage, on the basis of theoretical data for the dependence of the aberration coefficients and focal length on lens power. The total aberration shows a minimum value at a given value of the lens power, in a lens of fixed aperture and maximum field strength. The minimum does not occur at either the minimum focal length or the minimum of the spherical aberration coefficient, but at intermediate values of these quantities. When the maximum field is 10,000 oersteds, the minimum resolution is 10-12 A. at an accelerating voltage of 50 kv., the corresponding focal length being 2 mm. If the aperture is adjusted, in a lens of otherwise fixed dimensions, to the optimum value at each voltage, the total aberration varies very little at high voltages. If the lens dimensions are appropriately increased as well as the aperture, the total aberration can be progressively reduced with increasing voltage.

---

§1. INTRODUCTION

ELECTRON lenses are subject to the same five aberrations as are optical lenses : astigmatism, coma, distortion, curvature of the field and spherical aberration. They also show chromatic aberration when the incident beam is not monochromatic. Magnetic electron lenses show three additional errors, arising from the rotation of the image in the focusing field.   In light microscopy the correction of lenses is so far advanced that the primary limitation on resolving power is the diffraction effect, which depends directly on the wave-length and inversely on the

* The chief conclusions of this paper were briefly reported to the Electron Microscope Conference held in Manchester on 16 January 1946.   Comments have now been added on the treatment of the problem in D. Gabor's *The Electron Microscope*, which has since become available.

aperture; large apertures are therefore employed. Magnetic lenses, on the other hand, are still in a primitive state of development and are almost entirely uncorrected. The form which has been evolved for electron microscopy, largely by trial and error, suffers from a high degree of spherical aberration, which depends on the third power of the aperture; this is the limiting factor on resolution, the other aberrations being at present negligible in comparison. It is therefore necessary to operate magnetic lenses with such minute apertures (10^{-2} to 10^{-3} radian) that the diffraction effect, negligible at large and intermediate apertures on account of the small wave-lengths of the electrons, becomes important and a working compromise has to be reached: it has been usual to employ that aperture size for which the spherical and diffraction errors are equal. In order to keep the chromatic aberration (proportional to the aperture) smaller than these errors, it is then necessary to provide a high degree of smoothing of the voltage supply.

The effect of variation in accelerating voltage on the resolution limit is complex. The diffraction error depends on the equivalent wave-length of the electron beam, and hence inversely on the root of the voltage; it is therefore *prima facie* advantageous to operate at high voltage. As regards spherical aberration, previous discussion (on lines already summarized (Cosslett, 1945)) has led to the conclusion that it can be continuously decreased by raising the accelerating voltage and by reducing the lens dimensions. The former conclusion rested on the application to strong lenses of an approximate relation for the aberration of weak lenses (Scherzer, 1933), where the spherical coefficient depends inversely on the voltage. A reduction in dimensions appears to be advantageous because the degree of spherical aberration (at given voltage and lens field strength) depends on the product of the focal length of the lens and this coefficient, which is dimensionless and characteristic of the shape of the field distribution. The focal length falls in proportion to the lens dimensions, having a minimum value in a given lens equal to the half width (a) of the distribution curve at half the maximum field strength (H_0). For the purpose of evaluating the theoretical resolving power of an electron microscope, the value of the product (coefficient × focal length = C_s) has sometimes been taken as unity, but more usually a minimum value for a given field has been calculated or estimated, under the assumption that optimum resolution occurred at minimum C_s, and that this minimum value could be attained in practice.

It seemed desirable, from such a treatment of the problem, to employ the maximum practicable voltage in electron microscopy, in order to achieve reduction in both diffraction and spherical errors. However, attempts to improve resolution by operating magnetic microscopes at very high voltages (200-300 kv.) did not yield the expected result, either in the U.S.A. or in Germany. There has been some discussion (Ardenne, 1942; Dosse, 1942) on the reasons for this failure, but with no agreed conclusion. Gabor (1945) has also given attention to the problem, but underestimates the effect of voltage variation, owing to his use of the weak lens approximation for the dependence of focal length on voltage, and overestimates the effect of varying the lens dimensions, owing to a misinterpretation of Ramberg's results. The present investigation shows the determining factors to be the variation of the spherical aberration coefficient with voltage

and the saturation limit to the field strength set by the properties of the iron in the poles of the lenses.

An electron lens differs from its optical counterpart in being a "dynamical" lens, its power being continuously variable by changes in accelerating voltage V and field strength H, as well as in its dimensions. Its aberrations are complicated functions of V, H and their derivatives with respect to axial distance, as well as of the aperture. In a lens of given dimensions and, therefore, of given field distribution, both the aberration coefficient and the effective focal length will depend upon its power, which is represented by an expression of the form $cd^2H_0^2/V$, where c is a constant and d is a characteristic dimension, which may be the diameter D of the pole pieces or the field-width parameter, a. In practice it is desirable to operate with the maximum possible value of the field strength at the centre of the lens (H_0), as determined by the saturation properties of the iron circuit, in order to employ maximum voltage at a given lens power and thus
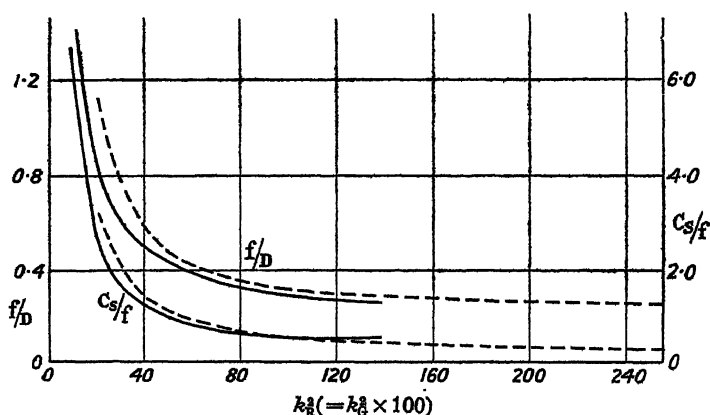
Figure 1.

For Glaser's field distribution, $H_z = H_0/(1 + z^2/a^2)$: $\left(k_G^2 = \dfrac{ea^2H_0^2}{8mV} \right)$.

For Ramberg's field distribution, $H_z = H_0 \operatorname{sech}^2 az$: $\left(k_R^2 = \dfrac{D^2H_0^2}{2 \cdot 63^2V} = 100k_G^2 \right)$.

minimize the diffraction error. Primary importance, therefore, attaches to the variation of the characteristics of a lens with voltage, at constant H_0.

Sufficient data are now available on the theoretical behaviour of lens aberrations with lens power (Glaser, 1941; Ramberg, 1942) to permit calculations of lens performance which take into account all the variables. It proves to be the case that the spherical aberration coefficient (C_s/f) falls with increasing lens power (figure 1), but at constant H_0 it must therefore increase with accelerating voltage, since power $\propto 1/V$. This increase is so rapid as to offset the parallel reduction in the diffraction error with voltage. The total aberration (assuming the chromatic error to be negligible) thus exhibits a flat minimum in relation to voltage (figure 2), for a given value of H_0, and this minimum does not occur at the minimum of either the focal length or the spherical aberration coefficient of the lens, as assumed by Dosse, Ardenne and others. If the magnetic field is increased, the minimum aberration occurs always at the same numerical value of the lens power; its absolute

value decreases if the voltage is correspondingly raised, maintaining the power of the lens constant at this optimum value. On the other hand, it now appears that reduction in lens dimensions offsets the favourable effect of voltage increase. Once the saturation field strength has been reached, reduction in dimensions demands a lower voltage for optimum resolution, a conclusion arrived at experimentally by Dosse (1941). Conversely, attempts to raise the working voltage with lenses of fixed dimensions and limited field strength can only result in a deterioration in resolution.

Once the saturation field strength has been reached, it is now shown that further improvement in resolution can only be attained by increasing simultaneously both the working voltage and the lens dimensions. The resolution improves in inverse proportion to the fourth root of the voltage if the dimensions
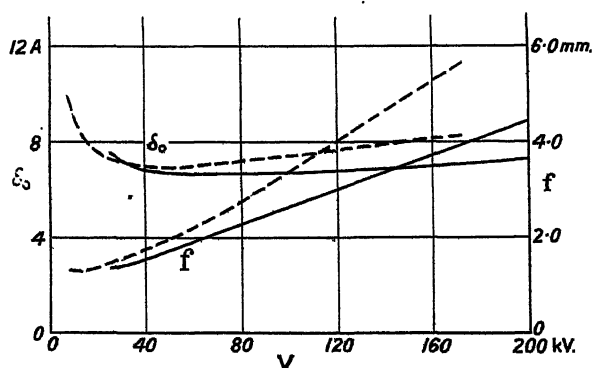
Figure 2.

Focal length, and minimum error ($\delta_0 = \delta_s = \delta_D$), as functions of accelerating voltage at fixed H_0 and D.   ($H_0 = 10{,}000$ oersteds ; $D = 4a = 5$ mm.).

Field distribution, $H_z = H_0/(1 + z^2/a^2)$: $-------$

,, ,, $H_z = H_0 \operatorname{sech}^2 az$: ————————

are increased in proportion to the root of the voltage. On this basis it is possible to design a magnetic electron microscope having a theoretical resolving power of a few Ångström units only, together with the high penetrating power required for the examination of many specimens, particularly of a biological nature.

The variation of chromatic aberration with lens power and voltage is also discussed. It shows a flat minimum value for a lens with fixed aperture size. If the aperture is varied to its optimum value for each lens power, as required by consideration of the spherical and diffraction errors, a very flat minimum appears, but at a power corresponding to a higher voltage at fixed field strength.

§2. SPHERICAL ABERRATION AND DIFFRACTION

The confusion in an electron microscope image is due primarily to diffraction and spherical aberration. The former effect depends inversely on the value of the aperture (α) and the latter directly on its third power. If δ_d and δ_s are the diameters of the respective discs of confusion, referred back to the object plane, then

$$\delta_d = K\lambda/\alpha \qquad \qquad \ldots\ldots(1\,a)$$

and

$$\delta_s = C_s\alpha^3, \qquad \qquad \ldots\ldots(1\,b)$$

where λ is the wave-length of the " illuminating " electron beam and C_8 is a spherical aberration coefficient characteristic of a given lens. K is a numerical constant dependent on the assumptions made as to when two diffraction discs become resolvable by the eye; we shall assume for the time being that it is unity, and discuss later the results of a more exact wave mechanical treatment (Glaser, 1943). Since the two errors vary in opposite senses with aperture, there is an optimum value of aperture giving minimum aberration. It is usual to assume that this occurs when the two errors are equal (Rebsch, 1938), so that

$$\delta_d = \delta_s = \delta_0 = \sqrt[4]{C_s\lambda^3} \quad \text{and} \quad \alpha_0 = \sqrt{\lambda/C_s} \qquad \ldots\ldots(2)$$

if $K = 1$. We prefer to use δ_0 here, for the optimum aberration at given λ, and to reserve $\delta_{\min}$ to designate its minimum value as λ is varied by changing the accelerating voltage, V, in a given lens.

From de Broglie's relation, and neglecting relativistic corrections, we have $\lambda = \sqrt{150/V}$ A., and, therefore, with C_s also in Ångström units,

$$\delta_0 = 150^{\frac{3}{8}} . \sqrt[4]{C_s V^{-\frac{3}{2}}} \quad \text{and} \quad \alpha_0 = 150^{\frac{1}{4}}/\sqrt[4]{C_s V^{\frac{1}{2}}}. \qquad \ldots\ldots(3)$$

If C_s were independent of voltage, then we should have

$$\delta_0 \propto V^{-\frac{3}{8}} \quad \text{and} \quad \alpha_0 \propto V^{-\frac{1}{4}}.$$

In fact, the value of C_s varies with voltage in a manner dependent on the form of the magnetic field and the position of the object in the lens. In a microscope objective the object is deeply immersed in the field, and for high magnification should be very close to the focal point. This discussion assumes the object to occupy such a position, varying with change in the power of the lens.

The variation of spherical aberration with lens power has been calculated by both Ramberg and Glaser for field distributions which approximate closely to those found in lenses of present design. The important quantity is C_s/f, which is dimensionless and characteristic of a particular form of field distribution, being invariant with respect to changes in lens dimensions providing that all are changed to scale. Glaser has calculated C_s/f for the field distribution $H_z = H_0 \Big/ \left(1 + \dfrac{z^2}{a^2}\right)$, where H_z is the axial field strength at the point z, H_0 is the maximum value of the field at the axis, and a is a lens parameter related to the half-width of the field distribution curve at $H_0/2$, and in this particular case equal to it. A rigorous solution of the electron path equation is possible with this field distribution, and an exact calculation of aberrations can be made. The power of the lens depends on the integral of H_z^2 taken along the axis from object to image, and, therefore, Glaser specifies the power of a lens in terms of a parameter $k^2 = \dfrac{ea^2H_0^2}{8mV}$. His results for C_s/f and the relative focal length $f/4a$ are given in figure 1 for the conditions of high magnification specified above.

Ramberg investigated the very similar distribution $H_z = H_0 \operatorname{sech}^2 ax$ by a ray-tracing and numerical integration method. He found the aberration in terms of a constant $C (= C_s/f)$ and specified the lens power by the quotient $D^2H_0^2/2\cdot63^2V$, where D is the minimum diameter of the pole pieces of the lens. Confusion has arisen in comparing his results with those of Glaser, which appear to give much

V. E. Cosslett

smaller values of the aberration coefficients. However, direct comparison is only legitimate if the two calculations are referred to lenses of the same dimensions and focal length. To this end we here assume $D = 4a = 5$ mm., and transform Ramberg's results into the set of parameters employed by Glaser; the results are plotted in figures 1 and 2. The two focal-length curves fit very well at a minimum value of $f/a = 1$, and the aberration curves also coincide satsfactorily; the degree of divergence is not more than might be expected from the difference in the field distributions operated upon in the two cases. As regards lens power, we find that the above assumption ($D = 4a$) leads to a simple relation between Ramberg's and Glaser's parameters: $D^2 H_0^2/2 \cdot 63^2 V = 100 k^2$, to a close approximation.

Both the aberration coefficient and the focal length curves fall with increasing lens power (k^2); that is, with increase of field strength, decrease of voltage or a degree of both together. The limiting factor in lens operation is the saturation
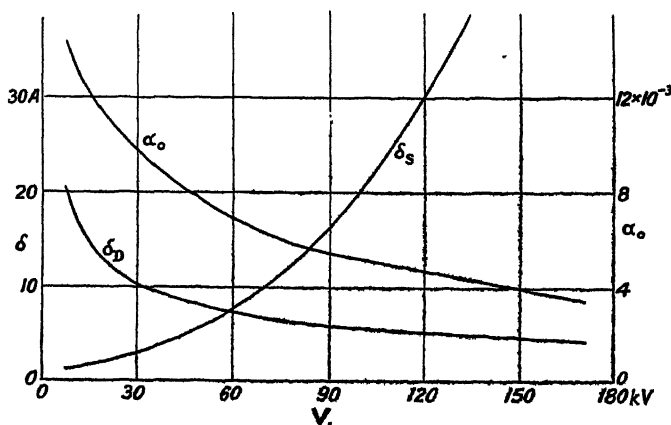
Figure 3.

Optimum aperture α_0 (for $\delta_s = \delta_d$) as a function of voltage, V. Spherical aberration δ_s and diffraction error δ_d as functions of V at *fixed* aperture ($a = 7 \cdot 10^{-3}$). (For $H_s = H_0/(1 + z^2/a^2)$; $H_0 = 10{,}000$ oersteds ; $a = 1 \cdot 25$ mm.)

of the pole pieces, and, since diffraction and other considerations make it desirable to use the highest possible voltage, it is usually necessary to work with field strengths close to saturation, in order to keep the lens power high and f and C_s/f small. We assume the saturation field to be 10,000 oersteds, and from this value of H_0 and the assumed value of a (or D), the value of the voltage corresponding to any value of k^2 can be calculated. Figure 1 allows C_s to be evaluated, and hence the value of δ_0 to be found for each voltage. Figure 2 shows the results thus calculated for both Glaser's and Ramberg's field distribution. That is, it gives the optimum aberration that can be obtained with a lens having the assumed dimensions (D) and saturation field (H_0). The two sets of results agree very well, showing a minimum aberration ($\delta_{min.}$) of 7 A. at an accelerating voltage of 60 kv.; but the minimum is flat and the aberration does not exceed 10 A. over a wide range of voltage. The corresponding focal length given by the data of figure 1 is also shown, indicating that the minimum value occurs at the comparatively low voltage of $11 \cdot 4$ kv.; the optimum f, corresponding to $\delta_{min.}$, is 2 mm. It is clear

that minimum aberration does not coincide with minimum focal length nor with minimum C_s, as has sometimes been assumed. The same data yield the optimum value of the aperture for each voltage, from equation (3). The curve (figure 3) shows a regular rise from about $3 \cdot 5$ to $14 \cdot 5 \times 10^{-3}$ with decreasing voltage, having a value of 7×10^{-3} at the voltage giving minimum aberration.

The results show, from the good agreement obtained from calculation on two different approximations to the actual field distribution, that a spherical aberration and diffraction error better than 10 A. is attainable by lenses of the present type, the values of D and H_0 taken being much the same as realized in practice. The combination of these errors with the chromatic aberration to give the total aberration is discussed later, but it may be remarked here that the optimum voltage and aperture found are very close, by accident or design, to those employed in present-day microscopes. The minimum aberration ($\delta_{min.}$) will always occur for the same value of k^2, as a (or D) and V are varied at constant H_0, but its absolute value will decrease as the voltage is raised. Since $V \propto a^2$ at constant k^2 and H_0, it follows that the limiting value of $\delta_{min.}$ can only be reduced below 7 A. by increasing both the voltage and the dimensions of the lens. Further consequences of this interrelation are reserved for §5.

§3. CHROMATIC ABERRATION

The diameter δ_c of the disc of confusion (referred to the object plane), due to chromatic aberration alone, is

$$\delta_c = \alpha \cdot \frac{\Delta V}{V} \cdot C_c, \qquad \ldots \ldots (4)$$

where $\Delta V/V$ is the fractional fluctuation in voltage and C_c is a coefficient depending on the field distribution and focal length of the lens. The voltage fluctuation is comprised of the residual ripple, $y \cdot V$, in the high-tension supply to the cathode, together with the thermionic variation in the energy of emission of electrons, x volts. So that

$$\Delta V/V = (x + y \cdot V)/V = x/V + y. \qquad \ldots \ldots (5)$$

Variation in the focusing field (ΔH) may be included in the factor x, for present purposes.

In a given lens, with constant focusing field H, the chromatic aberration varies with voltage owing to the change of all three variables, α, C_c and $\Delta V/V$. The effect of fluctuations will depend on the relative values of x and y, but only with very high smoothing efficiency (order of $2 \cdot 10^{-5}$) does the thermionic fluctuation become important, and then only at low voltages. For a given pole-form, and hence field distribution, the non-dimensional ratio C_c/f varies only slowly with lens power, and hence voltage. But the focal length depends on the voltage, at fixed H, and, therefore, the coefficient C_c increases with voltage in the usual range of operation of present lenses, the variation is approximately given by

$$C_c \propto V^{\frac{1}{2}}. \qquad \ldots \ldots (6)$$

The aperture α may be kept constant at a value designed for best average performance over a range of voltages, or it may be varied with the voltage, in which case it will affect the value of the chromatic aberration. Considering first the

aperture as fixed, substitution of equations (5) and (6) in equation (4) and differentiating shows that δ_c will have a minimum value when $V = x/y$. Taking Glaser's values for C_c/f for the field distribution $H = H_0\,(1 + z^2/a^2)$ and the lens conditions previously assumed ($H_0 = 10,000$, $a = 5/4$ mm.), we can calculate the variation of δ_c with voltage at a fixed value of the aperture (figure 4). The aperture chosen was the value (7×10^{-3}), corresponding to minimum spherical and diffraction error ($\delta_{\min}$) in such a lens. The curve shows a pronounced minimum of about 4 A., but lies below 10 A. over a very wide range of voltage (6–180 kv.). It rises sharply at lower voltages, where the thermionic fluctuation x becomes relatively important. In these calculations the value of the latter was taken as 1 v. and the ripple (y) as 1/40,000. The minimum value of δ_c is then predicted as occurring at $x/y = 40$ kv.; the fact that it occurs at about 33 kv. indicates the extent to which equation (6) is approximate.
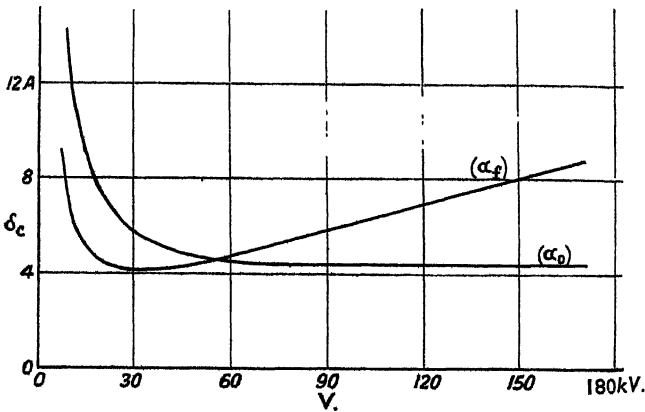
Figure 4.

Chromatic aberration δ_0 as function of voltage, at fixed aperture ($\alpha = 7 \cdot 10^{-3}$), and for variable aperture (α_0). Lens conditions as in figure 3.

In the second place we assume α to be adjusted at each voltage to the value (α_0), giving equality of spherical and diffraction errors, according to figure 3. The latter shows that $\alpha_0 \propto V^{-\frac{1}{4}}$ very closely. Inserting this relation, and equation (6) in equation (4), we have for the chromatic error

$$\delta_c = k \,.\, \Delta V/V = k(x/V + y)$$

from equation (5). Hence δ_c is determined in these circumstances solely by the relative values of the thermionic and ripple fluctuations, and should fall to a constant value dependent on y at high voltages, for which x/V is negligible. The curve calculated from Glaser's values of C_c for this case is given also in figure 4. It has the expected form, the indication of a very flat minimum at high voltage showing that the simplified relations assumed are only first approximations. At 57·3 kv. the two curves for δ_c intersect, this being the voltage for which the value of the aperture assumed for the first curve is equal to α_0 as defined. At lower voltages the equality of δ_s and δ_d requires a rapid rise in α and, therefore, in δ_0, so that the second curve lies above the first. At higher voltages, on the other hand, the value of α_0 falls below 7×10^{-3} and the chromatic error is lower for the second

·case considered. In both cases δ_0 is lower than δ_s and δ_d over the range of voltage from 20 to 150 kv.

§4. TOTAL ABERRATION

The values of δ_s, δ_d and δ_c obtained above have to be combined in order to obtain the total aberration observed. It is difficult to form a physical picture of their mutual effect; they are certainly not independent, and, therefore, simple addition is not justified. Both spherical and chromatic aberration will have the effect of producing a number of overlapping diffraction maxima in the image plane. It has been usual to assume, following Ardenne (1938), that they should be superimposed as error functions, and the geometrical mean taken as the total aberration. Such a procedure probably errs on the side of pessimism, but may be provisionally adopted so long as observed aberrations are so much larger than those predicted. We thus have for the total aberration δ_t in the object plane,

$$\delta_t = \sqrt{\delta_s{}^2 + \delta_d{}^2 + \delta_c{}^2} = \sqrt{2\delta_0{}^2 + \delta_c{}^2}, \qquad \ldots\ldots(7)$$

when optimum aperture is employed.

The application of this synthesis to the results given above for the individual aberrations yields the curves of figure 5 for the variation in δ_t with voltage both
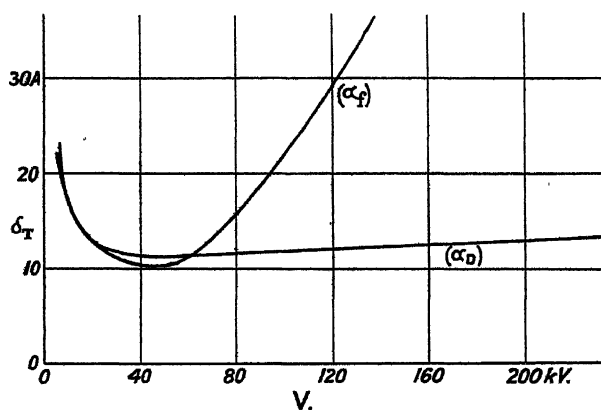
Figure 5.

Total aberration δ_T as function of voltage, at fixed aperture ($a = 7 \cdot 10^{-3}$), and for variable aperture (a_0). Lens conditions as for figures 3 and 4.

for a fixed aperture and for variable optimum aperture α_0. For the latter condition the total aberration initially falls rapidly with voltage, but then lies between 11 and 13 A. over the range 20 to 170 kv., with a flat minimum at about 50 kv. For a fixed aperture the minimum aberration naturally occurs near the voltage (57·3 kv.) for which α was fixed, as giving minimum δ_0; the increase in δ_s with higher voltage moves the minimum back to about 48 kv. For the same reason the total aberration curve for fixed aperture rapidly rises above that for optimum aperture at high voltages, where $\alpha > \alpha_0$, and, therefore, spherical aberration is large.

The above procedure has assumed that the combination of spherical and diffraction errors has a minimum when the two are equal. It was on this assumption that equation (2) was derived and equation (3) employed to find the combined

aberration when the aperture is α_0, as defined in terms of this equality; so that, neglecting δ_c, we have $\delta_t = \sqrt{2}\delta_0 = 1\cdot41 \sqrt[4]{C_s\lambda^3}$. But this is only a convenient working rule, having no theoretical basis, and the same is true of the importation of the diffraction expression directly from optics.

Gabor (1945, chapter 7) assumes that the minimum resolution occurs at the true minimum value of the sum of the squares of δ_d and δ_s, found by inserting their analytical relations (equations (1 a) and (1 b), above) and differentiating. Minimum resolution is then found when $\delta_d = \sqrt{3}\,.\,\delta_s$, whence $\delta_0 = 1\cdot33 \sqrt[4]{C_s\lambda^3}$. But he also assumes a more favourable value for the diffraction error, justifiable on physiological grounds of density discrimination in the eye: $\delta_d = 0\cdot82\lambda/\alpha$, in place of relation 1 a. His working expressions therefore become, with C_s in cm.,

$$\delta = 1\cdot14\sqrt[4]{C_s\lambda^3} = 750\sqrt[4]{C_s V^{-\frac{3}{2}}} \quad \text{and} \quad \alpha_0 = \sqrt[4]{\lambda/\sqrt{3}\,.\,C_s},$$

not sensibly different from the values used in the above calculations (neglecting δ_c):

$$\delta_t = 1\cdot41\sqrt[4]{C_s\lambda^3} = 925\sqrt[4]{C_s V^{-\frac{3}{2}}} \quad \text{and} \quad \alpha_0 = \sqrt[4]{\lambda/C_s}.$$

Gabor, however, proceeds to make the weak-lens assumption that f is proportional to accelerating voltage and to neglect the variation of C_s/f with voltage. He thus finds $\delta_{min} \propto V^{-\frac{1}{4}}$ in a given lens, showing a very slow fall with voltage, in place of the increase at high voltages demonstrated here (figures 2 and 5).

Glaser (1943) has recently given a rigorous wave mechanical treatment of the problem, in which diffraction and spherical aberration are treated together as one phenomenon, arising from the focusing of a portion only of a non-spherical wave-front. He obtains an expression for δ_0 similar in form to equation (2), but with a numerical factor

$$\delta_0 = 0\cdot56\sqrt[4]{C_s\lambda^3}. \qquad \ldots\ldots(8)$$

In terms of the above approach, this result may be interpreted in the (extreme) senses that either the diffraction or the spherical error is smaller than assumed. If the former, we may write $\delta = 0\cdot42\lambda/\alpha$ in place of equation (1 a), leaving the spherical aberration equation (1 b) unaltered. Assuming minimum error to occur when the two are equal, we arrive at Glaser's equation (8); the practical effect would be that an aperture of $0\cdot8\,\alpha_0$ would give a value of δ_0 smaller by $0\cdot56$ than that calculated from equation (3). On the other hand, we may write $\delta_s = 1/16\,.\,C_s\alpha^3$, and assume that the diffraction equation remains exactly as equation (1 a), in order to obtain equation (8). In this case the practical conclusion is that an aperture of $2\alpha_0$ will give the predicted smaller value of δ_0. If, as is probable, neither of these limiting cases is valid, we shall have numerical factors in equations (1 a) and (1 b) larger than $0\cdot42$ and $1/16$ respectively, giving a working aperture rather larger than α_0 for a combined error δ_0 smaller by the factor $0\cdot56$.,

Glaser himself gives a value for the optimum aperture, $\alpha_0 = 1\cdot13\sqrt[4]{\lambda/C_s}$, which yields a diffraction error only 5% smaller than the combined error (equation (8)). This is on his own assumption that the optical diffraction expression applies, $\delta_d = 0\cdot6\lambda/\alpha$; any other less favourable diffraction expression, such as Gabor's, or equation (1 a) of this paper, would lead to a diffraction error greater than Glaser's "total" aberration. Since the spherical aberration must play an appreciable part in determining the total confusion in the electron microscope

image, Glaser's conclusions must be regarded as of doubtful application. Their derivation is in any case insecurely based, owing to a lack of accurate information as to the details of the scattering of electrons in the object.

The effect of Glaser's result would be that the theoretical minimum values of the total aberration as here calculated and shown in figure 5 are reduced approximately by one-half. Or alternatively, the same calculated aberration would be obtained with the use of a larger aperture. Since it is not clear exactly how Glaser's result is to be interpreted in its effect on the value of the optimum working aperture, and since other aberrations (including the varying rotation of the image due to voltage fluctuations) have not been taken into account, it has been felt advisable to adhere to the classical method of treatment. The synthesis of aberrations obtained must therefore not be regarded as anything more than an approximate result, erring probably on the high side.

§ 5. DISCUSSION

The results of the present calculations show that there are definite optimum conditions of operation for a given magnetic electron lens. Its dimensions and maximum attainable field strength being fixed, there are particular values of the aperture and operating voltage, and hence of the focal length, which will give minimum total aberration. For present electron microscopes the conditions have been rather fortuitously fulfilled, owing to the flat minimum in the curve of total aberration against voltage even for a fixed value of the aperture. The conditions of operation are therefore not at all critical so long as the working resolution is of the order of 50 to 100 A., much greater than the theoretical limit, on account of limitations in materials and workmanship. But now that research microscopes are becoming capable of resolutions of 20 to 25 A. (Hillier and Baker, 1945), attention to such optimum considerations is more and more essential. The lens data assumed in the above aberration calculations are typical of lenses now in use. Figure 5 shows that their best theoretical performance is 10–11 A. for the total aberration for an aperture of 7×10^{-3} and at a voltage of 50 kv. Practical results are therefore already no more than double those given by theory.*

Attempts have been made in the past to improve the resolving power by operating at very high voltage, up to 300 kv., relying upon a supposed reciprocal relation between voltage and aberration. It is now clear that failure was due to the saturation limit to the focusing field strength, as it was not appreciated that aberrations increase rapidly if the field is not increased in proportion to the root of the voltage—that is, if the power of the lens is not kept constant; this is shown by figure 5. Assuming that the lens is of fixed dimensions, including aperture, and operating at maximum H_0, the total aberration rises rapidly with voltage above a value of about 60 kv. At 172 kv. it is already 67 A., the effective power of the lens having fallen to 0·2 from the optimum value of 0·6.

Since, for a given lens, there is an optimum value of the power (k^2) which gives minimum δ_d and optimum α, it follows from the definition of k^2 that as the working voltage is varied the magnetic field should be varied according to $V^{\frac{1}{2}}$, in order to keep the power constant at this value. From equation (3) it then follows

* *Note added in proof*: Hillier (1946) has now reported the achievement in favourable circumstances of a resolution of 10 A.

that the value of $\delta_{min.} \propto H^{-\frac{3}{4}}$. The value of δ_c also falls with increasing field strength and voltage, at optimum focal length, to a degree which depends on whether α is constant or always made equal to the optimum value corresponding to $\delta_d = \delta_s$. The table shows the variation in the several aberrations with field strength for the optimum conditions found above: $k^2 = 0.6$, $f = 2.1$ mm., when $a = 1.25$ mm. $\approx D/4$. As in figure 4, δ_c becomes preponderant at low fields for $\alpha = \alpha_0$, owing to the rapid increase in the latter, whereas for fixed aperture its value is less, whilst δ_d increases, owing to the increase in wave-length of the beam not being offset by an increase in aperture; δ_s is constant in this latter case. It will be seen that the net result gives very much the same total aberration for both fixed and variable aperture at all except very low field strengths. It should be mentioned that this discussion assumes constant high magnification, the focal length being constant throughout at its optimum value.

Table

Variation in individual and total aberrations with field strength H_0 (and corresponding acceleration voltage V) for a lens operated at optimum power $(k_0^2 = 0.6)$. $f_0 = 2.1$ mm.; $a = 1.25$ mm.; $H_z = H_0/(1 + z^2/a^2)$. $\delta_t = \sqrt{\delta_s^2 + \delta_d^2 + \delta_c^2} = \sqrt{2\delta_{min.}^2 + \delta_c^2}$.

H_0 (oers.)	V (kv.)	$\delta_{min.}$ (A.)	α_0 (rad.)	δ_c (A.)	δ_t (A.)	At fixed aperture, $a = 7 \cdot 10^{-3}$			
						δ_c (A.)	δ_d (A.)	δ_s (A.)	δ_t (A.)
1,000	0·57	40·3	$12·7 \times 10^{-3}$	334	339	185	72·5	7·2	199
2,000	2·29	23·9	10·7	73·1	80·5	48·7	36·3	7·2	61·0
4,000	9·17	14·2	9·0	17·9	26·9	14·1	18·1	7·2	23·9
6,000	20·6	10·5	8·1	8·8	17·3	7·7	12·1	7·2	15·8
8,000	36·7	8·5	7·6	5·9	13·3	5·5	8·1	7·2	12·7
10,000	57·3	7·2	7·2	4·5	11·4	4·7	7·2	7·2	10·8
12,000	82·5	6·2	6·8	3·8	9·6	4·0	6·0	7·2	10·0
15,000	129	5·3	6·5	3·1	8·1	3·5	4·8	7·2	9·2

The necessary conditions for design and operation of high-voltage lenses of high resolution follow from the above discussion, and are further considered in a subsequent paper. The expression for the lens power, $k^2 = ea^2H_0^2/8mV$, shows that the power and focal length can be kept constant at their optimum values, as the voltage is raised, by increasing a and H_0. Once the saturation field strength is reached, the remaining variable is a, the field width parameter. Hence, to ensure optimum resolution at high voltages, it is essential to increase the dimensions of the lenses, and not to decrease them, as has been the tendency empirically followed hitherto.

Gabor (1945, chapter 12) has noted, from the results of Ramberg, the fall in C_s/f with increase in lens dimensions at constant H_0 (cf. figure 1). He fails to observe, however, that it is f/D, and not f, which falls also with increasing lens power. If such an increase is brought about by changing the lens dimensions alone, at fixed field and voltage, it proves that f itself remains almost constant, the fall in f/D being roughly proportional to the increase in D, the lens diameter.

Hence only a small improvement in resolution results in such a case, due to the reduction in C_s. On the other hand, the above discussion has shown that better resolution is obtained if the lens power (and hence f/D) is kept constant by increasing both V and the lens dimensions. C_s/f also remains constant, whilst f increases. The increase in $\sqrt[4]{C_s}$, however, is more than offset by the effect of increasing the voltage, since $V^{-\frac{3}{4}}$ enters into the expression for δ_0 (equation (3)). It proves, in fact, that in these circumstances $\delta_0 \propto V^{-\frac{1}{2}}$.

If the maximum value of H_0 is 10,000 oersteds in the lens discussed earlier, then its dimensions will need to be increased some 2·4 times for operation at 300 kv. instead of at 57·3 kv., since at constant k^2 and H_0 we have $a \propto V^{\frac{1}{2}}$. Hence the pole diameter will need to be about $2·4 \times 5 = 12$ mm., and the optimum focal length will be approximately 5 mm. For the same magnification the microscope will need to be increased in length by the same factor, since the object distance is effectively the same as the focal length. Such an inconvenient increase in size can be avoided either by working at a lower electronic magnification, and employing subsequent photographic enlargement from fine-grain plates, or by resorting to multi-stage operation, with several lenses each giving small magnification. The latter solution appears to be most satisfactory if high voltage operation is required primarily for the higher resolution it makes possible, and not simply for the greater penetrating power of the beam, allowing thicker specimens to be observed at rather rough resolution.

REFERENCES

ARDENNE, 1938. *Z. Phys.*, **108**, 338 ; 1942. *Ibid.*, **118**, 384.
COSSLETT, 1945. *J. Sci. Instrum.*, **22**, 170.
DOSSE, 1941. *Z. Phys.*, **117**, 422; 1942. *Ibid.*, **118**, 375.
GABOR, 1945. *The Electron Microscope* (Hulton Press).
GLASER, 1941. *Z. Phys.*, **117**, 285; 1943. *Ibid.*, **121**, 647.
HILLIER, 1946. *J. Appl. Phys.*, **17**, 307.
HILLIER and BAKER, 1945. *J. Appl. Phys.*, **16**, 469.
RAMBERG, 1942. *J. Appl. Phys.*, **13**, 582.
REBSCH, 1938. *Ann. Phys. Lpz.*, **31**, 551.
SCHERZER, 1933. *Z. Phys.*, **80**, 193.

NEW REGULARITIES IN VIBRATIONAL SPECTRA

By K. M. GUGGENHEIMER,

H. H. Wills Physical Laboratory, University of Bristol

ABSTRACT. The frequencies of about 150 diatomic molecules are expressed as functions of the equilibrium distance and of the numbers of electrons in the outer shells of both atoms. The frequencies are calculated for homopolar molecules with single bonds, with multiple bonds and for molecules with strong polarity. The contributions of s- and p-electrons to the force constant are practically equal, and the same parameters are valid for all periods of the periodic system. The theoretical implications are discussed.

§1. INTRODUCTION

THE properties of molecules can be derived in quantum theory from first principles. Exact solutions have been calculated for those molecules which are composed of the lightest atoms. For molecules, however, which contain atoms with many electrons, the mathematical difficulties have not yet been overcome. It is therefore of importance and of interest to find empirical rules which correlate properties like frequencies, force constants, dissociation energies and interatomic distances. The following discussion is particularly concerned with the vibration frequencies of diatomic molecules in the ground state.

Various regularities have already been suggested. Kratzer (1920) found for the halogen hydrides the relation

$$\omega^2 I = \text{const.,} \qquad \qquad \ldots\ldots(1)$$

where ω is the wave number and I the moment of inertia. Birge (1925) and Mecke (1925) proposed for other molecules a relation

$$\omega I = \text{const.} \qquad \qquad \ldots\ldots(2)$$

Morse (1929) showed that the frequencies of various molecules are related by

$$\omega r_e^3 = \text{const.,} \qquad \qquad \ldots\ldots(3)$$

where r_e is the equilibrium distance.

Badger (1934) gave a formula for the force constants

$$k(r_e - d_{ij})^3 = 1\cdot86 . 10^5 \text{ dyne/cm.} \qquad \qquad \ldots\ldots(4)$$

The values d_{ij} vary according to the period in the natural system of elements to which the atoms of the molecule belong, from $d_{ij} = 0\cdot025$ for H_2 to $d_{ij} = 1\cdot635$ for I_2. The force constant k is related to the frequency of the molecular vibration ω by the formula

$$k = 4\pi^2\mu c^2\omega^2 N^{-1} = 5\cdot889 . 10^{-2}\mu\omega^2 \text{ dyne/cm.,} \qquad \ldots\ldots(5)$$

where μ is the reduced mass of the molecule in units of atomic weight and N the Avogadro number.

Allen and Longair (1935) consider

$$kr_e{}^6 = \text{const.} \qquad \ldots\ldots(6)$$

as an adequate formula. Here again the constant varies according to the period to which both atoms belong.

Fox and Martin (1939) find that in some organic molecules

$$kr_e{}^5 = \text{const.} \qquad \ldots\ldots(7)$$

gives better results.

The formulae (4), (6) and (7) are certainly superior to (2) and (3), since if one compares molecules in which one atom is replaced by an isotope, the equilibrium distance is not altered and also the force constant remains unchanged, but the vibration frequency must be different owing to the different mass of the oscillator (cf. Herzfeld (1933), and Allen and Longair (1935)).

There remains, however, the difference of the forms (4) and (6), which is of theoretical importance.

§ 2. THE CHARGE DEPENDENCE

A further complication arises because the force constant may depend not only upon the atomic distance and upon the period, i.e. quantum number, but also upon the number of electrons or valency electrons of both atoms. Clark and co-workers (1941 *et seq.*) investigated in a series of papers whether the molecular vibration frequency is a function of the sum n of electrons of both atoms which do not belong to closed shells and came to the conclusion that

$$kr_e{}^6 n^{\frac{1}{2}} = \text{const.} \qquad \ldots\ldots(8)$$

and

$$kr_e{}^6 n = \text{const.} \qquad \ldots\ldots(9)$$

describe equally well the empirical material. Again the constants in (8) and (9) vary with the periods to which both atoms belong, e.g. in (8) from $0\cdot107$ to $9\cdot90$.

Linnett discusses another kind of charge dependence of the repulsive force in molecules, namely, the form

$$\left(\frac{z_1 + z_2}{z_1 z_2}\right)^{2/3}, \qquad \ldots\ldots(10)$$

where z_1 and z_2 are the numbers of electrons in the outer shells of the two atoms.

It is plausible to suppose that the interaction of the two atoms depends to a first approximation upon the outermost electrons only, but it seems strange that the force constant should diminish with increasing number of electrons, as is implied by formulae (8) and (9), and also by (10), if, for example, $z_1 = z_2$.

One would rather expect that the interaction, and therefore the force constant, would increase with the number of electrons involved.

If the force constant depends upon distance, and also upon number of electrons, it is obvious that one has to try to separate the two effects, either by comparing molecules with similar atomic distances or by comparing molecules whose atoms have the same number of outer electrons. The first method would give at once an indication of the form of the charge dependence. Incidentally, there are a few

molecules in which the distance between the two atoms is nearly the same, e.g.

Molecule	r (A.)	k (dyne/cm.)	z_1 and z_2	
Li_2	2·6723	2·546.10⁴	1	$\dfrac{k(I_2)}{k(Li_2)} = 6\cdot75$
I_2	2·667	1·718.10⁵	7	

Lithium belongs to the first group of the periodic system, and iodine to the seventh. Therefore, one would expect that the ratio of the force constants should be a simple function of the ratio of the numbers of the outer electrons, in this case 7.

Other examples are provided by the pairs :

Molecule	r (A.)	k (dyne/cm.)	z_1	z_2	
KH	2·2244	5·555.10⁴	1	1	$\dfrac{k(BaH)}{k(KH)} = 1\cdot446$
BaH	2·232	8·034.10⁴	2	1	
CaH	2·002	9·697.10⁴	2	1	$\dfrac{k(CaF)}{k(CaH)} = 2\cdot70$
CaF	2·018	2·618.10⁵	2	7	
BeH	1·343	2·246.10⁵	2	1	$\dfrac{k(BeF)}{k(BeH)} = 2\cdot56$
BeF	1·3636	5·759.10⁵	2	7	

Assuming that the equilibrium distances of the fluorides of strontium and of barium are also nearly equal to that of their hydrides, one can compare their force constants :

Molecule	r (A.)	k (dyne/cm.)	z_1	z_2	
SrH	2·145	8·474.10⁴	2	1	$\dfrac{k(SrF)}{k(SrH)} = 2\cdot72$
SrF		2·302.10⁵	2	7	
BaH	2·232	8·168.10⁴	2	1	$\dfrac{k(BaF)}{k(BaH)} = 2\cdot65$
BaF		2·166.10⁵	2	7	

These values for the ratio of the force constants are similar to those of the previous analogous compounds of the second group, independent of the period in the natural system. This regularity is interesting in itself. But it becomes even more significant if this ratio is considered as a function of the numbers of electrons. In all cases mentioned above the force constant can be described by assuming that it depends upon the number of the outermost electrons according to the formula

$$k \sim \sqrt{(z_1 . z_2)}. \qquad \dots\dots (11)$$

In fact, it gives for the four cases of fluorides and hydrides of the alkaline-earth metals a theoretical ratio of $\sqrt{7} = 2\cdot645$ compared with an empirical average of 2·66. The ratio of the sums $z_1 + z_2$ of the outermost electrons would be 3. In the case of BaH against KH formula (11) gives a value $\sqrt{2} = 1\cdot4141$ compared with 1·446 empirically, and for I_2 against Li_2 $\sqrt{49} = 7$ compared with 6·75 empirically.

In all cases the experimental deviation is only a few per cent. Further evidence for this formula will be found later on.

The relation (11) is rather surprising if compared with the Coulomb law. The possibility of a wave-mechanical foundation will be discussed later.

The dependence of the force constant upon the distance can now be determined. For this purpose those molecules will first be compared which have an analogous constitution, i.e. which consist of atoms of similar valency so that the number of outer electrons in corresponding atoms remains constant. The simplest cases are provided by the molecules of hydrogen and the alkali metals. Using the method of least squares for the series H_2, Li_2, Na_2 and K_2, whose distances have been measured with spectroscopic exactitude, one gets

$$k = 2 \cdot 738 \cdot 10^5 r_e^{-2 \cdot 46} \text{ dyne/cm. (if } r \text{ is in A.),} \qquad \ldots\ldots (12)$$

or, since the observed quantities are the frequency and the distance, the formula for the wave number is

$$\omega = 2 \cdot 156 \cdot 10^3 \mu^{-\frac{1}{2}} r_e^{-1 \cdot 23} \text{ cm.}^{-1} \text{ (} r \text{ in A.),} \qquad \ldots\ldots (13)$$

where μ is the reduced mass in units of atomic weight.

Combined with formula (11) this gives

$$\omega = 2 \cdot 156 \cdot 10^3 (z_1 \cdot z_2)^{\frac{1}{2}} \mu^{-\frac{1}{2}} r_e^{-1 \cdot 23} \text{ cm.}^{-1} \text{ (} r \text{ in A.)} \qquad \ldots\ldots (14)$$

as a general formula. Table 1a gives the values calculated according to (14) together with the distances and the observed frequencies. Those molecules are considered first whose intermolecular distance has been determined by the analysis of rotation spectra, since they represent the best measurements. Then molecules are taken whose interatomic distances have been measured by electron diffraction. The frequencies have been taken from G. Herzberg (1939), who gives a complete list of spectra measured up to 1939. Thomson and Cochrane (1939), and Maxwell (1940), give surveys of electron-diffraction measurements. In table 1 a, measurements of interatomic distances in polyatomic molecules have also been used on the assumption that the distance in diatomic molecules is not larger than the smallest separation of the two atoms found in polyatomic molecules with bonds of equal valency. In some cases more recent measurements by Lister and Sutton (1941) have been used. Furthermore, some molecules are inserted whose distances have been calculated by Pauling's radii (1940), which are based on x-ray crystal analysis and the additivity of covalent radii. Finally, some cases have been taken into consideration where use has been made of the fact that in the periodic system the radii of neighbouring atoms have in general similar values.

As can be seen from table 1a, not only the alkali metals and the halogens satisfy formula (14), but also many other group combinations, e.g. the hydrides of elements of groups not considered above, from boron to fluorine. Particularly interesting is the long series of molecules of atoms with 4 and 7 electrons. Again the power law (13) is fulfilled. Moreover, the absolute values of the frequencies confirm the charge dependence of (11) even if these molecules are not compared with molecules with similar equilibrium distance. The transition to other groups shows also that the influence of the charge is given correctly by (11). The contributions of s- and p-electrons to the force constant are practically equal.

It is remarkable that a simple formula like (14) comprises so many cases (64 in table 1a), though it contains only two constants. Moreover, since the hydrogen molecule whose properties can be calculated theoretically by wave

mechanics, is also included, there is really only one constant, the power of the distance, which has to be determined.

Table 1 *a.* Molecules with single bonds : $\omega = 2 \cdot 156 \cdot 10^3 (z_1 \cdot z_2)^{\frac{1}{2}} \mu^{-\frac{1}{2}} r_e^{-1 \cdot 23}$ (cm.$^{-1}$).

	r_e (A.)		$\omega_{obs.}$	$\omega_{calc.}$	Error	$z_1 z_2$	
H$_2$	0·7414	s	4405·3	4405·3	(standard)	1×1	
Li$_2$	2·6723	s	361·3	344·1	2·1 %	1×1	
Na$_2$	3·079	s	159·2	159·4	0·1	1×1	
K$_2$	3·923	s	92·64	90·9	1·9	1×1	
Cl$_2$	1·989	s	564·9	585·3	3·6	7×7	
Br$_2$	2·284	s	323·2	326·6	1·0	7×7	
I$_2$	2·667	s	214·4	214·2	0·1	7×7	
HF	0·9166	s	4141·3	4005	3·3	1×7	
HO	0·9710	s	3728	3606	3·3	1×6	
HN	1·038	s	(3300)	3180	3·6	1×5	
HC	1·1201	s	2824	2760	2·3	1×4	
HB	1·2326	s	(2366)	2291	3·2	1×3	
BiCl	2·48	e	307·7	306	1·6	5×7	
BiBr	2·63	e	209·3	211	0·8	5×7	
CCl	1·73	e	(843·6)	845	0·2	4×7	$^2\Pi$ (CHF$_2$Cl)
SiF	1·603	s	856·7	849	0·9	4×7	$^2\Pi$
SiCl	2·00	e	535·4	536	0·1	4×7	$^2\Pi$
SiBr	2·15	e	425·4	425	0·1	4×7	$^2\Pi$
GeCl	2·08	e	408·4	403	1·3	4×7	$^2\Pi$
GeBr	2·29	e	296·6	290	2·2	4×7	$^2\Pi$
SnCl	2·32	e	352·4	353	0·2	4×7	$^2\Pi$
SnBr	2·44	e	247·7	250	0·9	4×7	$^2\Pi$
PbCl	2·43	e	303·9	304	—	4×7	$^2\Pi$
PbBr	<2·60	e	207·5	203	2·2	4×7	$^2\Pi$
BCl	1·73	e	830·0	811	2·3	3×7	
BBr	1·87	e	686·3	687	0·1	3×7	
AlCl	2·06	e	481·3	485	0·8	3×7	
AlBr	2·21	e	379·2	387	2·1	3×7	
AlI	2·53	e	316·1	312	1·3	3×7	
HgCl	2·23	e	293·8	285	3·0	3×7	
HgBr	2·44	e	186·2	185	0·6	2×7	
ZnI	2·42	e	223·4	216	3·3	2×7	
CdI	2·56	e	178·5	170	4·8	2×7	
HgI	2·55	e	125·0	126	0·8	(1×7)	
BiI	2·79	P	163·9	167	1·9	5×7	
GaCl	2·25		365·0	352	3·6	3×7	Wells
GaBr	2·37		263·0	263	0·1	3×7	*r*
GaI	2·54		216·4	219	1·2	3×7	Ga 1·26
InCl	2·43		316·8	299	5·6	3×7	In 1·44
InBr	2·55		221·0	213	3·6	3×7	Tl 1·47
InI	2·72		177·1	173	2·6	3×7	
TlCl	2·46		287·5	280	2·6	3×7	Cl 0·99
TlBr	2·58		192·5	191	0·8	3×7	Br 1·11
TlI	2·75		150	150	—	3×7	I 1·28
ZnF	1·59		619·5	617	0·4	2×7	*r*(ZnH)
HgF	1·74		490·8	506	3·1	2×7	HgH
BiF	1·81		510·7	513	0·5	5×7	BiH
SbF·	1·82		614·0	612	0·8	5×7	TeF e

	r_e (A.)		$\omega_{obs.}$	$\omega_{calc.}$	Error	$z_1 z_2$	
SnF	1·82		585·3	584	0·2	4×7	TeF(SnO)
PbF	2·01		507·2	508	0·2	4×7	PbO r_P
NBr	1·91		693	686	1·0	5×7	CBr e
LiK	3·30		(207)	204	1·5	1×1	r
LiRb	3·42		(185)	187	1·0	1×1	Li 1·34
LiCs	3·55		(167)	176	5·4	1×1	Na 1·54
NaK	3·50		123·3	121	1·9	1×1	K 1·96
NaRb	3·63		106·6	104	2·4	1×1	Rb 2·09
NaCs	3·75		(97)	98	1·0	1×1	Cs 2·21
Rb$_2$	4·18		56·78	57	0·4	1×1	$\frac{1}{2}r_e$
RbCs	4·30		49·41	49·8	0·8	1×1	Cl$_2$ 0·99
Cs$_2$	4·42		41·99	42·5	1·2	1×1	Br$_2$ 1·14
IBr	2·48		268·4	268	0·1	7×7	I$_2$ 1·33
BrCl	2·14		(430)	456	6·0	7×7	
ICl	2·32	s	384·2	390	1·3	7×7	

s = spectroscopic
e = electron diffraction
P = Pauling radii.

At the end of table 1 *a* the radii of rubidium and caesium have been computed by using the known radii of the hydrides and by assuming that the difference of the distance in the hydrides and the radius of the alkali atom is the same as that of potassium. The calculated frequencies compare well with the observed ones, and formula (14) thus appears to be applicable from H_2 to Cs_2, i.e. from the smallest to the largest stable diatomic molecules.

§3. MULTIPLE BONDS

So far, only molecules with a single bond have been considered. Table 1 *b* gives the data for molecules with double bonds. The force constant varies again as $r^{-2\cdot46}$, as in molecules with single bonds described in equation (12). One new constant comes in indicating the strength of the double bond. Formula (14) is replaced by

$$\omega = 2\cdot976 \,.\, 10^3 \,(z_1\,.\,z_2)^{\frac{1}{2}}\mu^{-\frac{1}{2}}r_e^{-1\cdot23}\,\text{cm}^{-1} \qquad \ldots\ldots(15)$$

The first factor at the right-hand side is increased by a factor 1·38. If the double bond contributed a factor 2 to the force constant, the further factor for (15) would be 1·414. The empirical factor 1·38 means that the double bond contributes a factor 1·905 to the force constant.

Table 1 *b* again gives distances, and observed and calculated wave numbers. First those molecules are listed whose interatomic distances have been determined spectroscopically. Then a series of molecules is considered whose interatomic distances are derived from Pauling's radii on the assumption of the additivity of covalent radii. It is remarkable how good the agreement is between calculated and observed wave numbers in the long series of molecules with $z_1 = 4$ and $z_2 = 6$, even where the distances have not been observed directly. Among these molecules are also those measured recently and discussed by Barrow (1944) and Jevons (1944).

The series with molecules of the nitrogen type can also be described by the same formula which has been derived from molecules which possess double bonds

from the chemical point of view. The former are supposed to possess a triple bond. There is, however, no appreciable change in the relation between force constant and equilibrium distance. In this respect they belong to the same bond order.

Table 1 *b*. Molecules with multiple bonds: $\omega = 2 \cdot 976 \cdot 10^3 (z_1 \cdot z_2)^{\frac{1}{2}} \mu^{-\frac{1}{2}} r_e^{-1 \cdot 23}$ (cm.$^{-1}$)

	r_e (A.)	$\omega_{obs.}$	$\omega_{calc.}$	Error	$z_1 z_2$	
CO	1·1284 s	2168	2168	(standard)	4×6	
SiO	1·510 s	1242	1243	0·1 %	4×6	
TiO	1·620 s	1008	1050	4·2	4×6	$^3\Pi_u$
GeO	1·651 s	985·6	980	0·6	4×6	
SnO	1·838 s	822·4	829·9	0·9	4×6	
CS	1·563 s	1285·1	1301	1·2	4×6	
PbS	2·395 s	428·1	422·5	1·3	4×6	
N_2	1·095 s	2359·6	2251	4·6	5×5	
P_2	1·898 s	780·4	773·1	0·9	5×5	
NP	1·491 s	1337 ·	1312	1·9	5×5	
CN	1·172 s	2068·7	2036	1·6	4×5	
CP	1·5622 s	1240	1237	0·2	4×5	
SiN	1·572 s	1152	1181	2·5	4×5	
BO	1·205 s	1885·4	1907	1·2	3×6	
NO	1·150 s	1907	1940	1·7	4×5	$^2\Pi$
PO ·	1·447 s	1231	1230	0·1	4×5	$^2\Pi$
NS	1·50	1220	1226	0·5	4×5	$^2\Pi$ r(SO)
AsO	1·61	960·6	964	0·3	4×5	$^2\Pi$ (SeO)
CSe	1·735 P	1036	1036	—	4×6	
SiS	2·01 P	749·5	726	3·1	4×6	$r_{Pauling}$
SiSe	2·14 P	580·0	569	2·0	4×6	C 0·665
SiTe	2·34 P	480·4	483	0·5	4×6	Si 1·07
GeS	2·06 P	575·8	576	—	4×6	Ge 1·12
GeSe	2·19 P	406·8	406	0·2	4×6	Sn 1·30
GeTe	2·39 P	323·4	329	1·7	4×6	Pb 1·46
SnSe	2·37 P	332·0	329	0·9	4×6	
SnTe	2·57 P	259·5	262	1·0	4×6	O 0·55
SnS	2·24 P	487·7	487	0·1	4×6	S 0·94
PbO	2·01 P	721·8	724	0·3	4×6	Se 1·07
PbSe	2·53 P	277·4	277	0·1	4×6	Te 1·27
PbTe	2·73 P	211·8	216	2·0	4×6	
						r
As_2	2·19	429·4	414	3·5	5×5	(GeSe)
Sb_2	2·57	272·8	270	1·0	5×5	(SnTe)
Bi_2	2·92 P	173·7	174	0·2	5×5	
AsN	1·65	1068	1049	1·8	5×5	(GeO)
SbBi	2·73	220	221	0·6	5×5	(PbTe)

In this series the frequencies of some molecules are computed according to (15) on the assumption that the distance of, for example, Sb_2 is the same as in SeTe, the sum of the radii of the preceding and of the following elements.

All these molecules with multiple bonds obey the same power law, $k \sim r^{-2 \cdot 46}$, as molecules with single bonds.

§4. POLAR MOLECULES

There is another class of molecules ; they do not obey the $r^{-2\cdot46}$ relation. Most hydrides, and also the halides of the alkali and of the alkaline-earth metals, belong to this class. It is easy to find a simple formula for these molecules also, again with only two constants, of the same form as (14), i.e. of the type $k = Cr^s$.

Formula (14) has to be replaced by

$$\omega = 2\cdot065 \,.\, 10^3 \,(z_1\,.\,z_2)_{\frac14}\mu^{-\frac12}r_e^{-0\cdot92}\,\text{cm.}^{-1}\ (r\text{ in }\text{A.}).\quad \dots\dots\dots(16)$$

The values for the two constants have been derived by using NaH and CsH as standard molecules. The vibrations of KH and RbH are represented by the same formula. Moreover, not only BaH but also the other hydrides of the alkaline-earth metals obey formula (16) and confirm the charge dependence of formula (11) within the experimental error. It is worth noticing that the vibrations of alkali halides can also be described by the same formula, taking $z_1 = 1$; $z_2 = 7$.

One important conclusion can be drawn from this fact : according to (16), the wave number is a function of the product of the numbers of outer electrons. From the agreement of calculated and observed wave numbers one can conclude that the molecules are built up not by ions but by neutral atoms. If this is true even for the alkali halides, it is also true for the other molecules.

Most of the molecules whose frequencies obey formula (16) are known to have a strong dipole moment. This can be considered as responsible for the fact that the force constant of these molecules varies much more slowly with distance than that of molecules whose dipole moment is small or zero.

The group of molecules of the oxygen type belongs also to this series. They are known to have a $^3\Sigma$ ground state. The parallelism of the spins seems to be sufficient to create a contribution to the force constant.

In table 2 all known hydrides of metals have been listed for the sake of completeness. As in table 1, the numbers of outer electrons used for the calculations of the frequencies correspond to the maximum valencies of the atoms. However, the frequencies of the hydrides of Cu, Ag and Au, which belong to the transition elements, are much higher than expected from equation (16). This is probably due to the effect of the last attached d-electrons at small distances. One could use equation (16) to determine an effective number of electrons which allows the best description of the empirical frequencies in these three cases. They are shown in brackets in table 2. The halides of copper behave normally. .

Table 2. $\omega = 2\cdot065 \,.\, 10^3 \,(z_1\,.\,z_2)^{\frac14}\mu^{-\frac12}r_e^{-0\cdot92}\,(\text{cm.}^{-1})$

	r (A.)	$\omega_{\text{obs.}}$	$\omega_{\text{calc.}}$	Error	$z_1\,z_2$
HNa	1·8875 s	1170·8	1170·8	(standard)	
HCs	2·494 s	890·7	890·7	(standard)	
HK	2·244 s	983·3	989·8	0·7 %	1 × 1
HRb	2·368 s	936	935·8	—	1 × 1
HLi	1·5956 s	1406	1430	1·7	1 × 1
HBe	1·343 s	2058	1965	4·5	1 × 2
HMg	1·7302 s	1495	1506	0·7	1 × 2
HCa	2·002 s	1299	1307	0·6	1 × 2
HSr	2·1457 s	1207	1218	0·9	1 × 2

Table 2—*continued*

	r (A.)		$\omega_{obs.}$	$\omega_{calc.}$	Error	$z_1 z_2$	
HBa	2·232	s	1172	1173	0·1	1×2	
HZn	1·5947	s	1608	1604	0·2	1×2	
HCd	1·762	s	1431	1458	1·9	1×2	
HHg	1·741	s	1387	1472	6·1	1×2	
HTl	1·870	s	1391	1525	9·6	1×3	
HPb	1·839	s	1565	1664	6·3	1×4	
HBi	1·809	s	1699	1786	5·1	1×5	
HAl	1·6461	s	1683	1744	3·6	1×3	
HSi	1·521	s	(2080)	2109	1·4	1×4	
HP	1·433	s	(2380)	2244	5·7	1×5	
HCo	(1·543)	s	(1890)	1831	3·1	1×3	
HNi	1·475	s	(1927)	1904	1·2	1×3	
HCu	1·463	s	1940	1921	1·0	(1×3)	
HAg	1·618	s	1760	1785	1·4	(1×3)	
HAu	1·5239	s	2306	2275	1·3	(1×7)	
HCl	1·2747	s	2980	2713	9·0	1×7	
HBr	1·414	s	2650	2444	7·8	1×7	
HI	1·604	s	2310	2174	5·9	1×7	
O_2	1·2076	s	1580	1497	5·3	6×6	$^3\Sigma$
S_2	1·89	e	725·8	701	3·4	6×6	$^3\Sigma$
Se_2	2·16	e	392·5	392	0·1	6×6	$^3\Sigma$
Te_2	2·59	e	251·5	262	4·2	6×6	$^3\Sigma$
SO	1·4935	s	1123·7	1066	5·1	6×6	$^3\Sigma$
SeO	1·61	e	908·9	891	2·0	6×6	$^3\Sigma$
TeO	1·82		796·1	768	3·5	6×6	
NaCl	2·51	e	380	385	1·3	1×7	
NaBr	2·64	e	315	325	3·2	1×7	
NaI	2·90	e	286	285	0·5	1×7	
KCl	2·79	e	280	303	8·2	1×7	
KBr	2·94	e	231	242	4·8	1×7	
KI	3·23	e	212	208	1·9	1×7	
BeF	1·362	s	1266	1211	4·3	2×7	
CaF	2·018	s	587·3	579·4	1·3	2×7	r of
SrF	2·146		500·1	498	0·4	2×7	SrH
BaF	2·23		468·9	465	0·8	2×7	BaH
MgCl	<2·51		466·0	452	3·0	2×7	NaCl
MgBr	~2·64		373·2	379	1·6	2×7	NaBr
CaCl	<2·79		364·5	358	1·8	2×7	KCl
SrCl	~2·79		301·1	309	2·6	2×7	KCl
BaCl	~2·79		279·2	292	4·6	2×7	KCl
CaBr	~2·94		(280·2)	292	4·2	2×7	KBr
SrBr	~2·94		(212)	232	(9·3)	2×7	KBr
BaBr	~2·94		(192)	204	(6·1)	2×7	KBr
CuF	1·759	s	619·5	619·2	0·1	2×7	
CuCl	2·13	e	417·0	418	0·2	2×7	
CuBr	2·25	e	314·1	319	1·6	2×7	
CuI	2·40	e	264·8	268	1·2	2×7	
C_2	1·3121	s	1641·7	1600	2·5	3×3	$^3\Pi_u$

At the end of table 2, the vibrations of a few molecules have been considered whose intermolecular distances have not been measured directly but have been

inferred indirectly from molecules of similar size. The agreement between calculated and empirical wave numbers is sufficient to show that these molecules also belong to the series of polar molecules which obey formula (16).

There remains the possibility that there are also molecules with a double bond which belong to this series with the smaller power law. The molecule C_2 appears to be of this type. This molecule is known to have as ground state a $^3\Pi$ configuration. In molecules with triplet states the eigenfunction must be antisymmetric and, therefore, have a node at the centre of the line which goes through both nuclei. This means that in C_2 one pair of charges plays practically no rôle in a region where normally the interaction of the two atoms is greatest. Thus it is understandable that only three electrons of each atom are effective, and on this assumption (allowing for a factor 2 for the force constant of the double bond), the frequency of C_2 is given satisfactorily by (16).

A similar phenomenon seems to occur in the four molecules NO, PO, AsO and NS of table 1 b. All of them have doublet ground states ($^2\Pi$). This can mean that all electrons apart from the odd one form together either a singlet or a triplet state. The latter possibility would again give rise to a practical non-participation of a pair of electrons in the critical region between the two atoms. Empirically these four molecules show among themselves the correct dependence $r_e^{-1.23}$ upon the distance, but the charge product seems to be diminished by just one unit in each atom. This phenomenon seems to occur only with Π states with their pronounced directional property.

§5. DISCUSSION

In the foregoing it has been shown that for a substantial part of all molecules whose frequencies in the ground state have been measured, the force constant can be described by a simple function of the equilibrium distance. The formulae depend only on two constants. The difference between the first constant for the frequencies in formulae (14) and (16) is only 5%. The $r^{-1.23}$ power relation is fulfilled by about one hundred molecules ; the power rule $r^{-0.92}$ by about fifty.

The $r^{-1.23}$ relation covers the range from the highest observed wave number, 4405 cm.$^{-1}$ (H_2), to the smallest wave number, 42 cm.$^{-1}$ (Cs_2). In the series of molecules with multiple bonds the same power law extends from the molecule BO to $^{209}_{83}Bi_2$, the heaviest stable diatomic molecule.

The difference between calculated and observed frequencies is for the one hundred molecules of table 1, on the average 1·6%, and often less than the uncertainty due to the errors in the measurements of the equilibrium distance.

There is practically no difference between the effect of s- and p-electrons, nor is the period of the elements of any importance.

The dependence upon the distance is smaller than would be predicted by formulae (4), (6) or (7).

The fact that the variations of long series of molecules can be described by formulae with only two constants is not only remarkable in itself but must indicate a simple but fundamental physical fact.

It is understandable that the force constant between two atoms depends, to a first approximation, only on the number of electrons which are in

the outermost shells of both atoms. Ample evidence for, and confirmation of, formula (11) have been found in tables 1 and 2. The form of (11), the proportionality of the force constant to the geometric mean number of electrons, is, however, new and surprising.

In the following it will be shown how this relation can be understood within the framework of quantum mechanics. In wave mechanics the density of the charge, or the probability of finding one electron, is given by

$$\rho = \psi\psi^* d\tau \qquad \dots\dots (17)$$

and the eigenfunction is normalized by

$$\int \psi\psi^* d\tau = 1. \qquad \dots\dots (18)$$

If there were z_1 electrons with the same eigenfunction, the probability of finding one electron at a certain place would be z_1 times as great and, therefore,

$$\int \psi\psi^* d\tau = z_1. \qquad \dots\dots (19)$$

The Pauli principle does not allow more than one electron to be in a given state. The electrons, however, which belong to the same shell, i.e. have the same principal quantum number, have similar radial eigenfunctions, at least the s- and p-electrons. In particular the mean distance from the nucleus is rather similar. The probability of finding one electron in the outermost shell will therefore be z_1 times as great as if there were only one electron there. In fact, the method of the self-consistent field of Hartree is based on this condition.

Generally, the function which represents the solution for the whole system of electrons can be considered to be the result of the superposition of the eigenfunctions of all the electrons which belong to the system

$$\Psi_1 = \Sigma c_i \psi_i, \qquad c_i^2 = 1. \qquad \dots\dots (20)$$

As a matter of fact, the electrons are indistinguishable. In a system of electrons, each electron can belong to each state which is occupied. The probability of finding one electron at a certain position is equal to the sum of the probabilities due to the eigenfunctions which belong to the occupied states

$$\Psi_1 \Psi_1^* d\tau = \Sigma \psi_i \psi_i^* d\tau. \qquad \dots\dots (21)$$

If the eigenfunctions are orthogonal, condition (19),

$$\int \Psi_1 \Psi_1^* d\tau = \int \Sigma \psi_i \psi_i^* d\tau = z_1, \qquad \dots\dots (22)$$

is fulfilled, and the orthogonalization is always possible. The condition of the form (22) corresponds to the quantization of the amplitudes introduced by Dirac (1927) in the theory of radiation.

If there are two atoms, equation (22) has to be fulfilled also for the second atom,

$$\int \Psi_2 \Psi_2^* d\tau = z_2. \qquad \dots\dots (23)$$

If two separated atoms approach one another, the two outer shells will begin to overlap. This gives rise to a perturbation. The energy of the whole system becomes a function of the distance of the two nuclei. The force constant

is equal to the second derivative of the energy with respect to the interatomic distance, taken at the equilibrium distance.

According to equation (11) the force constant is empirically proportional to the geometric mean of the numbers of the outermost electrons,

$$k \sim \sqrt{(z_1 . z_2)}. \qquad \ldots \ldots (24)$$

With regard to equations (22) and (23), this fact obtains a simple theoretical meaning :—The force constant is proportional to the mixed density of the electrons

$$k \sim |\Psi_1 \Psi_2|. \qquad \ldots \ldots (25)$$

In the one-electron problem, H_2^+, the binding energy, is essentially due to the overlap integral $\int \frac{1}{r_1} \psi_1 \psi_2 \, d\tau$, if use is made of the method of linear combination of atomic orbitals, and also the force constant is, at least to a first approximation, theoretically proportional to the mixed density. The result of (25) appears as a natural extension of this fact.

The repulsive force contributes more to the magnitude of the force constant, since it varies more quickly with the distance than the force of attraction. At the equilibrium distance the repulsive force and the attractive force are equal in magnitude. The force constant is due to the difference of the derivatives of these forces. One cannot derive the law of one of these forces alone from the rules connecting force constant and distance given by equations (14) to (16), but the variation of the force constant with distance gives at least a lower limit for the variation of the repulsive force, and conversely, a given law for the repulsive force should give an upper limit for the variations of the force constant.

The most important contribution to the repulsive force is due to the pressure of the electronic gas which is built up by the electrons of both atoms, particularly of those which participate in the binding. To get an indication of the magnitude of these forces one has to consider the electronic gas confined to the volume of the molecule. The energy of the electronic gas is then, according to Fermi,

$$E = \frac{3}{40} \frac{h^2}{m} \left(\frac{3}{\pi} \right)^{2/3} \frac{z^{5/3}}{V^{2/3}}. \qquad \ldots \ldots (26)$$

The contribution to the force constant is determined by the second derivative with respect to the atomic distance. The order of magnitude is given by $\hbar^2/mr^4$. This yields for $r = 1$ A. an order of magnitude of 10^5 dyne/cm. This is necessary and sufficient to satisfy the requirements of equation (12), which describes the empirical facts.

The occurrence of the mass of the electrons in the expression $\hbar^2/mr^4$ shows that the electrons are responsible for the magnitude of the force constant, though the frequencies which are measured are due to the vibrations of the total masses of the atoms of the molecule.

The variation of the atomic distance during vibration leads to a variation of the volume of the electronic gas proportional to the atomic distance. As a result of equation (26) the variation of the energy is then proportional only to $r^{-2/3}$, where r is the distance between the two atoms, and taking the second derivative leads to $r^{-2 \cdot 67}$ for the variation of the force constant. Owing to the effect of binding forces

and of the non-existence of rigid limits for the electronic gas, the actual variation of the force constant will be slower. Comparison with the empirical formula $r^{-2\cdot46}$ of equation (12) shows that this consideration leads at least to a good approximation and yields the right order of magnitude for the exponent of the distance r in equation (12).

The relations for the force constant of diatomic molecules which have been derived in this paper can be applied also to polyatomic molecules, and can be compared with other properties of matter, e.g. the compressibilities and thermal expansion of solid bodies. This will be done in later papers.

REFERENCES

ALLEN and LONGAIR, 1935. *Phil. Mag.*, **19**, 1032.
BADGER, 1934. *J. Chem. Phys.*, **2**, 128.
BARROW, 1944. *Proc. Phys. Soc.*, **56**, 204.
BIRGE, 1925. *Phys. Rev.*, **25**, 240.
CLARK and WEBB, 1941. *Trans. Faraday Soc.*, **37**, 293.
DIRAC, 1927. *Proc. Roy. Soc.*, **114**, 243.
FOX and MARTIN, 1939. *J. Chem. Soc.*, 884.
HERZBERG, 1939. *Molecular Spectra and Molecular Structure.*
HERZFELD, 1933. *Handb. d. Phys.*, XXIV/2, 98.
JEVONS, 1944. *Proc. Phys. Soc.*, **56**, 211.
KRATZER, 1920. *Z. Phys.*, **3**, 289.
LINNETT, 1942. *Trans. Faraday Soc.*, **38**, 1.
LISTER and SUTTON, 1941. *Trans. Faraday Soc.*, **37**, 393, 406.
MAXWELL, 1940. *J. Opt. Soc. Amer.*, **30**, 375.
MECKE, 1925. *Z. Phys.*, **32**, 823.
MORSE, 1929. *Phys. Rev.*, **34**, 57.
PAULING, 1940. *Nature of the Chemical Bond.*
SKINNER and SUTTON, 1944. *Trans. Faraday Soc.*, **40**, 164.
THOMSON and COCHRANE, 1939. *Theory and Practice of Electron Diffraction.*
WELLS, 1944. *Structural Inorganic Chemistry.*

TIDAL EFFECTS ON THE PRODUCTION OF MESONS IN THE ATMOSPHERE

By A. W. MAILVAGANAM,
Colombo, Ceylon

ABSTRACT. The theory of forced oscillations, magnified by resonance, of the earth's atmosphere with a period of 12 solar hours, as worked out by Pekeris, is applied to calculate the amplitude of oscillation of the layer in the atmosphere in which mesons are produced by primary incident particles. The calculations show that, on account of these oscillations and of the finite life-time of mesons, a semi-diurnal variation in meson intensity with an amplitude of the order of 0·14 % may be expected.

THE main features of the semi-diurnal barometric oscillation attributed to tides in the atmosphere are now well known. It has two components, a solar and a lunar, the amplitude of the former being about 15 times that of the latter. To explain the solar component, Kelvin made the suggestion that the atmosphere has a free period of nearly 12 hours, so that the solar semi-diurnal tide is amplified by resonance. The free and forced oscillations of the atmosphere have been

studied by G. I. Taylor (1936) and by C. L. Pekeris (1937, 1939). Taylor has shown from a study of the propagation of waves of explosion that the atmosphere has a free period of oscillation of about 10·5 hours. Pekeris, using the known distribution of temperature within the first 60 kilometres of the atmosphere, and by assuming the simplest type of temperature distribution above 60 kilometres consistent with the existence of a free semi-diurnal oscillation in the atmosphere, has shown that the resulting free oscillation possesses the character· istics required by the "dynamo" theory of the diurnal variation of the earth's magnetic field. It also followed from Pekeris' work, in agreement with Taylor's results, that the atmosphere so "adjusted" possesses yet another period of free oscillation of a period of about 10·5 hours.

In an appendix to his second paper, Pekeris has given curves showing the variation of the semi-amplitudes of the vertical displacements of a particle with height above the ground for the two modes of oscillation. We note from the curve for the 12-hour period that these amplitudes become appreciable at altitudes above 10 kilometres. It may be recalled in this connection that Appleton and Weekes (1939), using radio methods of upper-atmospheric exploration, have detected a lunar tide in the Kennelly-Heaviside layer. They found in the experimentally determined expression for the equivalent height of the lower boundary of this layer a simple harmonical term, $0.93 \sin(2t' + 112)$ km. They concluded that the tide is semi-diurnal, in phase with the lunar barometric oscillations as observed by Chapman at Greenwich, and of the order of 1 kilometre.

Since, according to Pekeris, the amplitude of the oscillations of a particle increases rapidly with altitude, it is of interest to examine the extent to which the layer in which mesons are produced in the atmosphere oscillates owing to the semi-diurnal solar tide. For it is known from the work of Euler and Heisenberg (1938) that mesons are radioactively unstable, with a mean time of decay of about 2×10^{-6} sec. in a system of co-ordinates in which they are at rest. In consequence, if the layer of production rises in the atmosphere, then fewer mesons would reach the earth, owing to the increased path the mesons have now to travel. Blackett (1938) has successfully used these considerations to explain (i) the temperature effect of cosmic rays, and (ii) the observations of Auger, Ehrenfest, Freon and Fournier (1937), who found that the absorption of inclined cosmic rays was greater than that of vertical rays under the same thickness of absorber. Blackett showed that both effects could be explained in terms of meson decay, and deduced from the available data that the mean free path (decay) of the mesons in the atmosphere is about 24 km.

In this work account will be taken only of the solar semi-diurnal tide, the corresponding lunar tide being much smaller in comparison. The work of Pekeris (1937) will be followed closely and the notation employed by him used. In order to simplify the calculations the following stratification of the atmosphere, slightly different from that employed by Pekeris, is assumed. A rough check indicates that no appreciable error is introduced by this simplification.

The numbers 0, 1, 2, 3, 4 denote the boundaries between the sections of the atmosphere, 0 referring to the ground level. From the work of Hamilton, Heitler and Peng (1943) we take the meson-producing layer as lying between equivalent water-thicknesses of 15 and 30 cm. below the top of the atmosphere.

In centimetres of mercury these will correspond to 1·1 to 2·2. The summer heights in the atmosphere corresponding to these pressures are, according to Humphreys (1929), 29 and 24·5 km. respectively. In order to estimate the order of magnitude of the rise of the layer of meson production it would be sufficient to use either the summer or the winter values. It will be seen that this layer lies in the stratosphere (i.e. in region 1–2).

	Layer	Position in atmosphere (km.)	Temp. (° K.)	Remarks
Troposphere	0–1	0–10	287–230	Negative lapse rate
Stratosphere	1–2	10–37·5	230	Constant temp.
Layer of abnormal reflection of sound waves	2–3	37·5–60	230–270	Positive lapse rate
Adjusted top of atmosphere	3–4	60–80	270–215	Negative lapse rate
	4	>80	215	Constant temp.

For our purpose it is necessary to calculate the amplitude of the semi-diurnal tidal oscillation of the isobars passing through the heights under reference. The oscillations of any particular level have been calculated by Pekeris and represented graphically (figure 3, curve Z_2, of his paper (1939)). The magnitudes of the two kinds of oscillation must necessarily be different, the former being smaller owing to the heaping-up of the atmosphere under tidal influence. Using the notation employed by Pekeris, we have, if p is the pressure variation and h the displacement of an isobar under tidal influence,

$$\frac{\partial p}{\partial t} = \sigma p, \quad p = -g\rho h,$$

where $\sigma = 2\omega$; ω is the angular velocity of the earth's rotation, ρ the density of air at the level concerned, and g the acceleration due to gravity. These two equations, along with equation 34 (Pekeris, 1937),

$$\frac{1}{\rho}\frac{\partial p}{\partial t} = -\frac{\gamma}{2}gH\{C(1+v)e^{\lambda_1 y} + D(1-v)e^{\lambda_2 y}\},$$

yield

$$h = \frac{\gamma}{2\sigma}H\{C(1+v)e^{\lambda_1 y} + D(1-v)e^{\lambda_2 y}\},$$

where y is given by the relation $\left(\frac{287}{5\cdot7} - Z\right) \times 10^5$, Z being the height of a layer measured in kilometres, and γ is the ratio of the two specific heats of air. In the derivation of these equations we have omitted writing the imaginary quantity i, which only takes into account the phase of the oscillation. Furthermore, we have assumed ρ, the density of the air, to be constant over the distance h and equal to the value on the isobar concerned. The constants appearing in this equation have been calculated by the method indicated by Pekeris (1939, appendix), and are as follows:—

$$C = 0\cdot0794 \times 10^{-5}, \quad D = -0\cdot975 \times 10^{-5}$$
$$\lambda_1 = -0\cdot0634 \times 10^{-5}, \quad \lambda_2 = -0\cdot0851 \times 10^{-5}$$
$$V = 0\cdot1483, \quad H = 7\cdot87 \times 10^5 \text{ cm.}$$

Using these values, we find that the limits of the production layer estimated by Hamilton, Heitler and Peng are raised by $4 \cdot 12 \times 10^3$ cm. and $2 \cdot 76 \times 10^3$ cm. respectively at tidal maximum. Comparison with the curve given by Pekeris shows that these displacements of the isobars are smaller than, but of the same order of magnitude as, the displacements of layers at the respective heights.

From the above calculations we find that the lower surface of the production layer is raised at tidal maximum by $3 \cdot 44 \times 10^3$ cm. (mean), the mean height of the surface being 27 km. Using Blackett's estimate of the mean free path (decay) of the meson, we obtain the following equations for the mesons reaching the earth at zero tide and maximum tide respectively:—

$$N = N_0 e^{-2 \cdot 7/2 \cdot 4}, \quad N' = N_0 e^{-2 \cdot 7034/2 \cdot 4},$$

where N_0 is the number of mesons produced in the layer. We thus get as a lower limit for the percentage variation in meson intensity at tidal maximum

$$\frac{N - N'}{N} \times 100 = 0 \cdot 14 \%.$$

According to Simpson (1918), the semi-diurnal pressure maxima due to the solar tide occur at 10 a.m. and 10 p.m. local time. The minima in the meson intensity due to solar tides should therefore occur at these times. The tidal theory developed by Pekeris may be considered satisfactory in many respects; but it fails to explain why the semi-diurnal maxima occur at 10 a.m. and 10 p.m. (local time).* It may also be objected that the assumed temperature variation for the top of the atmosphere, which is by no means unique, does not fit in with the data obtained by the radio-exploration of the upper atmosphere.

In conclusion I would take this opportunity to thank Mr. M. Kidnapillai, of the Ceylon Technical College, for much valuable assistance in the calculations, and Mr. V. Appapillai for helpful discussions.

REFERENCES

APPLETON, E. V. and WEEKES, K., 1939. *Proc. Roy. Soc.*, A, 171, 171.
AUGER, P., EHRENFEST, P., FREON, A. and FOURNIER, A., 1937. *C.R. Acad. Sci., Paris*, 204, 257.
BLACKETT, P. M. S., 1938. *Phys. Rev.*, 54, 973.
BLACKETT, P. M. S., 1938, *Nature, Lond.*, 142, 992.
CHAPMAN, S., 1924. *J. Roy. Met. Soc.*, 50, 165.
EULER, H. and HEISENBERG, W., 1938. *Ergeb. d. Exakt. Nat.*, 17, 1.
HAMILTON, J., HEITLER, W. and PENG, H. W., 1943. *Phys. Rev.*, 64, 78.
HUMPHREYS, W. J., 1929. *Physics of the Air* (London), Table III.
PEKERIS, C. L., 1937. *Proc. Roy. Soc.*, A, 158, 650.
PEKERIS, C. L., 1939. *Proc. Roy. Soc.*, A, 171, 434.
SIMPSON, G. C., 1918. *Quart. J. Roy. Met. Soc.*, 44, 1.
TAYLOR, G. I., 1936. *Proc. Roy. Soc.*, A, 156, 318.
ROSSI, HEBERRY and HOAG, 1940. *Phys. Rev.*, 57, 461.

* This has, however, been explained by Chapman (1924).

NOTE ON MEASUREMENTS OF GLASS ABSORPTION

By T. SMITH, F.R.S.,

National Physical Laboratory

ABSTRACT. The theory of four methods of measuring the transmissivity of optical glass is given, with special reference to the effect of surface reflexion. The methods include one recently described by A. Maréchal, and two which have been used for industrial measurements. The accuracy attained by all four methods is similar provided the appropriate conditions, which are more stringent in some cases than others, are satisfied.

ONE of the chief requirements of glass from which optical instruments are to be made is a high transmission factor. Great attention is given to the purification of the raw materials, particularly important being a very low iron content in the sand. Correspondingly important is the accuracy of the measurements from which the transmissivity of the glass is derived.

The loss of light in passing through glass is attributable to two distinct causes. Within the glass, light is absorbed, and at the surfaces some is reflected. The surface losses depend on the state of the surface; a freshly polished surface may differ from a clean surface that has been exposed to the air for some days, and a bloomed surface will, in general, differ appreciably from an untreated surface. For these and other reasons it is desired to separate the two effects, and particularly to determine the true loss due to absorption within the glass, from which the transmissivity, i.e. the proportion of light transmitted through a thickness of 1 cm. of glass, is obtained. To obtain accurate values the specimens examined should give a long glass path, so that any errors in the measurements will have a very small effect on the value found for the transmissivity. For example, if the transmittance of a specimen 20 cm. long is 80 ± 1 per cent, the transmissivity is $0 \cdot 9889 \pm 0 \cdot 0006$, so that the accuracy is of the order of one part in two thousand, though the direct measurements are uncertain to about one per cent.

Before considering methods of measurement it is convenient to establish the relations between the glass and surface properties and the measurable quantities. At each surface some of the incident light is reflected and some transmitted, but it is assumed that none is absorbed; let the ratio of the reflected to the transmitted light be ϵ. Within the glass the fraction τ is transmitted and none is reflected.

The measurable quantities are the intensities of the beams of light entering and leaving the specimen. It is convenient to consider light travelling in both directions on each side of the specimen. Let the light intensities of the entering and leaving beams on one side be denoted by i and j respectively, and the beams on the other side travelling in the same directions be denoted by i' and j' respectively. The emergent light i' is composed of the transmitted part of i and the reflected part of j', and similarly j is composed of the reflected part of i and the

transmitted part of j'. Let t and r denote the transmission and reflexion factors of the complete assembly. It is known that t is independent of the direction in which the light travels, and that the values of r for both directions are equal if the system is symmetrical or when there is no absorption. The relations between the beams may be written concisely

$$(i'\ j')\begin{pmatrix} 1 & . \\ -r & t \end{pmatrix} = (i\ j)\begin{pmatrix} t & -r \\ . & 1 \end{pmatrix},$$

so that

$$(i'\ j') = (i\ j)\begin{bmatrix} t - \dfrac{r^2}{t} & -\dfrac{r}{t} \\ \dfrac{r}{t} & \dfrac{1}{t} \end{bmatrix}. \qquad \ldots\ldots(1)$$

When light falls on a succession of media the last equation shows that the resultant effect is obtained by multiplying in their proper order matrices of the form shown on the extreme right of this equation. If we note that when there is no absorption $t + r = 1$, and that this matrix then takes the form

$$\begin{pmatrix} 1 - \epsilon & -\epsilon \\ \epsilon & 1 + \epsilon \end{pmatrix}.$$

we see that for the glass specimen under consideration ·

$$\begin{bmatrix} t - \dfrac{r^2}{t} & -\dfrac{r}{t} \\ \dfrac{r}{t} & \dfrac{1}{t} \end{bmatrix} = \begin{bmatrix} 1 - \epsilon & -\epsilon \\ \epsilon & 1 + \epsilon \end{bmatrix}\begin{bmatrix} \tau & . \\ . & \dfrac{1}{\tau} \end{bmatrix}\begin{bmatrix} 1 - \epsilon & -\epsilon \\ \epsilon & 1 + \epsilon \end{bmatrix}. \qquad \ldots\ldots(2)$$

Clearly to obtain the transmittance τ two quantities must be measured.

When the refractive index μ of the glass is known, and the surfaces are freshly polished, ϵ has the value $\dfrac{(\mu - 1)^2}{4\mu}$. From the value of $\dfrac{1}{t}$ in equation (2) we obtain the quadratic equation

$$\tau = t\{(1 + \epsilon)^2 - (\tau\epsilon)^2\} \qquad \ldots\ldots(3)$$

· for τ. The percentage accuracy in τ is substantially equal to that in t. The experimental procedure is to determine the ratio of i' to i when $j' = 0$. In practice it is sufficient to know the type of glass being measured, as this will determine ϵ with ample accuracy provided a freshly polished specimen is used. If the surfaces are well bloomed ϵ will be small enough to be neglected.

A method that has been extensively employed is to make transmission measurements on two specimens of the same glass, one very thick and the other thin. Assuming that their surfaces are in a similar state, equation (3) gives, with ample accuracy,

$$\frac{\tau_1}{\tau_2} = \frac{t_1}{t_2} \cdot \frac{1 - \{t_1\epsilon(1 + \epsilon)\}^2}{1 - \{t_2\epsilon(1 + \epsilon)\}^2},$$

where a very rough value of ϵ is adequate. The left side of this equation represents the transmission factor for a thickness of glass equal to the difference of the

thicknesses of the two specimens. As the percentage errors in the values of t are algebraically subtractive, the corresponding error in τ may be twice that made in a measurement of t.

A third method that is less familiar is to determine both r and t experimentally. ϵ will disappear from equation (2) if we pre-multiply throughout by (1 1) and post-multiply by $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. This gives

$$\frac{(1-r)^2}{t} - t = \frac{1}{\tau} - \tau,$$

a quadratic equation for τ. When t and r are known, τ can therefore be found whatever the condition of the surfaces may be so long as there is no surface absorption. The two surfaces may be dissimilar or non-uniform in their properties, though in the latter event the same areas must receive the light in all measurements. If the errors in the measurements of t and r are δt and δr, the corresponding error in τ is $\delta \tau$ where,

$$\left(\frac{1}{\tau^2} + 1\right)\delta\tau = \left\{\left(\frac{1-r}{t}\right)^2 + 1\right\}\delta t + 2\frac{1-r}{t}\delta r,$$

or the error in τ is approximately the sum of those in t and r. As r tends to be small, the error in its value will normally be of little importance, and the error in τ is substantially equal to that in t.

A new method has been proposed recently by Maréchal (1944). He measures the brightness of the reflected beam when the glass specimen is backed by a mirror having a reflexion factor ρ. If R is the observed reflexion factor of the combined system, equation (1) gives

$$(1 \quad R) \begin{bmatrix} t - \dfrac{r^2}{t} & -\dfrac{r}{t} \\ \dfrac{r}{t} & \dfrac{1}{t} \end{bmatrix} \begin{bmatrix} -\rho \\ 1 \end{bmatrix} = 0$$

or

$$\frac{R-r}{\rho} = \frac{t^2}{1-r\rho},$$

since $j' = \rho i'$. When ϵ, and therefore also r, is small enough to be neglected the equation reduces to

$$\tau^2 = \frac{R}{\rho}.$$

This equation will be applicable when the surfaces of the specimen are carefully bloomed. The relative accuracy of the value of τ is equal to that of a single reflexion factor.

When ϵ is not negligible, substitution for r and t from (2) leads to a quadratic equation in τ^2 which may be written

$$\tau^2\rho =$$

$$[(1+\epsilon)\{1+\epsilon(1-\rho)\}-\tau^2\epsilon\{\rho+\epsilon(1-\rho)\}][(1+\epsilon)\{R-\epsilon(1-R)\}-\tau^2\epsilon\{1-\epsilon(1-R)\}],$$

so that τ^2 is a symmetrical function of R and $1/\rho$. In both factors the coefficients of ϵ tend to be small and those of ϵ^2 decidedly smaller, so that the correction

required owing to surface reflexion is substantially represented by first-order terms. This gives

$$\tau^2 = \frac{R}{\rho}\left[1 - \epsilon\left\{\frac{(1-R)^2}{R} + \frac{(1-\rho)^2}{\rho}\right\}\right] \qquad \ldots\ldots(4)$$

as a working formula.

The accuracy attained may be illustrated by a numerical example. Suppose the true values are

$$\epsilon = 0\cdot05, \quad \tau = 0\cdot08, \quad \rho = 0\cdot90.$$

Calculation by (2) gives

$$R = 0\cdot5851, \quad \frac{R}{\rho} = 0\cdot6501, \quad \frac{(1-R)^2}{R} + \frac{(1-\rho)^2}{\rho} = 0\cdot3053.$$

Substituting in (4) and using the correct value of ϵ gives

$$\tau^2 = 0\cdot6402,$$

a very satisfactory result. The magnitude of the difference between τ^2 and R/ρ indicates that ϵ should be known with an accuracy of a few units in the third decimal place. If the surfaces are bloomed the term in ϵ should be negligible.

ACKNOWLEDGMENT

REFERENCE

MARECHAL, A., 1944. "La mesure directe du facteur d'absorption des verres optiques." Paris: *Communications des Laboratoires de l'Institut d'Optique.*, No. 14.

DEMONSTRATION OF A WATER-JET ANALOGUE OF THE REFLECTION KLYSTRON

By W. J. SCOTT,

B.T.H. Research Laboratory, Rugby

Demonstrated 8 June 1945

THE reflection klystron has a rhumbatron-type cavity resonator with two closely spaced grids through which an initially unmodulated electron beam passes. The field between the grids alternates, accelerating or retarding the electrons, which pass on into a retarding field which returns them through the grids. The accelerated electrons go further than the retarded ones and so take longer to return. With proper adjustment of beam voltage and reflection field, the resonator receives over one half-cycle more energy than it gives up.

It is often taught that electric current is like water flowing through a pipe, and that a triode valve is like a water-valve. Extending the analogy (Scott, 1944), an electron beam can be represented by a water jet and an electric

field by the earth's gravitational field. .On this basis a working model equivalent to a reflection oscillator has been made, using a balanced horizontal arm A mounted on two adjustable leaf-spring hinges B as the tunable "resonator" (figure 1). On one end of the arm, a close-mesh gauze "catcher" grid C is fixed, and behind it, but 10 in. distant, a nearly vertical nozzle D which is fed with water through the hollow arm from a flexible rubber tube E. Figures 1 to 4 show the model in oscillation. The photographs were taken using a B.T.H. electrical discharge photo-flash lamp set to give a flash duration of 100 micro-seconds.

Starting from rest, the random motion of the water as it breaks into globules produces sufficient agitation (corresponding to electron agitation) to initiate oscillation provided the jet velocity is correct. In action the jet is modulated in velocity by the vertical simple-harmonic motion of the nozzle. The water accelerated during the second half of the up-stroke and that retarded on the first half of the down-stroke bunch together and arrive back to hit and propel the gauze grid on its downward stroke. Thus over one half-cycle more energy is given to the system than is absorbed by friction, windage, and by water impinging on the upward stroke.

Figure 1 shows a compacted bunch, FG, half-way back, and at the top a still widely spread "future" bunch, HJ. In figure 2 the beam is moving down, and is about to be hit by the now horizontally disposed bunch KL. At the top the next bunch, MN, is compacting. In figure 3 the bunch OP is splashing through the grid. In figure 4 the grid is moving up again, the next bunch, QR, forms a loop inclined at 45°, and a rising part, ST, of the jet, inclined at 10°, will form another. In figure 2 two well-formed bunches and another partly formed one are in the air.

Oscillations can be maintained by flight times corresponding to approximately $1\frac{1}{4}$, $2\frac{1}{4}$, $3\frac{1}{4}$, etc., cycles, instead of the $\frac{3}{4}$, $1\frac{1}{4}$, $2\frac{1}{4}$, etc., cycles, which are required by the electron tube. The slight increase or decrease of frequency, caused by slightly lowering or increasing the jet velocity, so that "bunches" arrive sooner or later at the catcher, corresponds strictly to the electronic tuning of reflex tubes (Pierce, 1945). If the jet velocity is definitely incorrect, oscillations will not build up or even maintain themselves.

REFERENCES

BRAINERD, KOEHLER, REICH and WOODRUFF, 1942. *Ultra High Frequency Techniques* (Chapman and Hall), p. 339.
British Patent 545,146.
HAHN, W. C. and METCALF, G. F., 1939. *Proc. Inst. Radio Engrs.*, 27, 109.
PIERCE, J. R., 1945. *Proc. Inst. Radio Engrs.*, 33, 112.
SCOTT, W. J., 1944. *J. Instn. Elect. Engrs.*, 91, 143.
VARIAN, R. H. and S. F., 1939. *J. Appl. Phys.*, 10, 140.
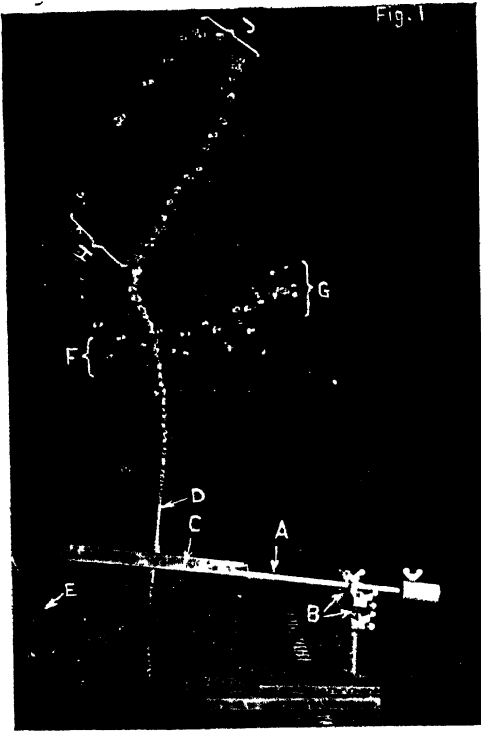
Figure 1.
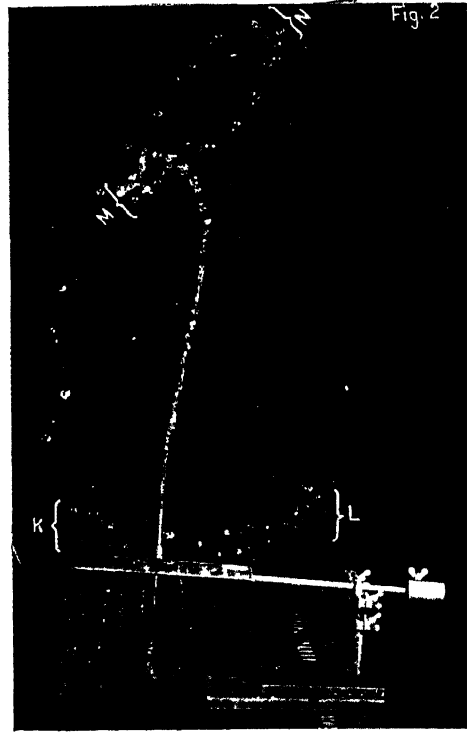
Figure 2.
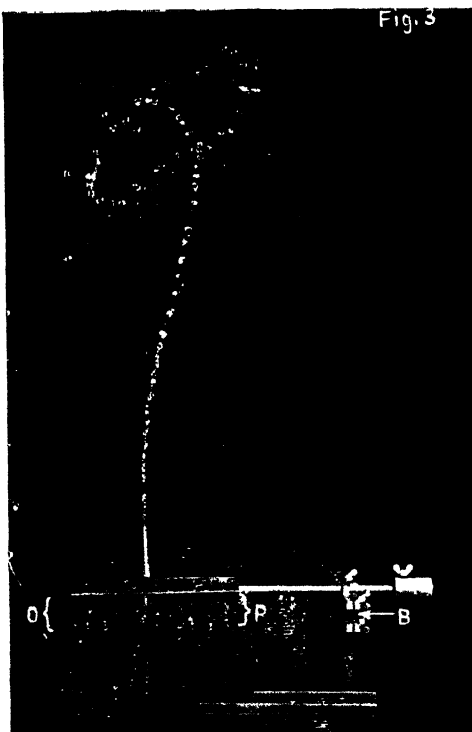
Figure 3.
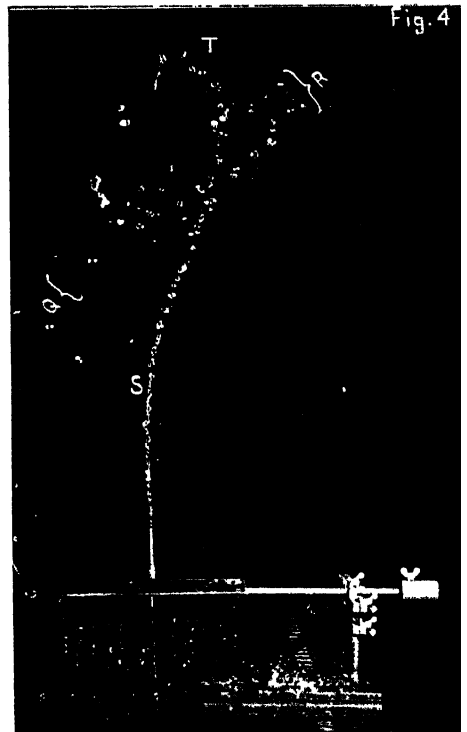
Figure 4.

NOTE ON THE MEASUREMENT OF SPECTRAL DISTRIBUTIONS WHEN LINES AND CONTINUA ARE PRESENT TOGETHER

By MARY P. LORD,

Philips' Lamps, Ltd.

Now at Imperial College, London

ABSTRACT. The problem of evaluating the intensity of a line spectrum relative to that of a continuum is solved for the specific case of a double monochromator '(the dispersions in the two halves of which are in the same direction) used in conjunction with a selective detector. The form of a function of the slit-widths and the dispersion required for the solution is derived analytically : a geometrical treatment given by Cittert is criticized.

§ 1. INTRODUCTION

As the radiation from the majority of fluorescent lamps consists of lines (mainly those of mercury) superimposed on phosphor continua, the problem of evaluating the intensity of a line spectrum relative to that of a continuum is one of practical importance. Despite this it appears to have received little attention.

§ 2. SOLUTION IN TERMS OF AN UNKNOWN FUNCTION OF THE SLIT-WIDTHS AND THE DISPERSION

To avoid circumlocution, the solution of the problem is given for the specific case of a double monochromator, the dispersions in the two halves of which are in the same direction (Perry, 1938) used in conjunction with a selective detector. It is further supposed that for each wave-length the energy of continuum per unit of wave-length (at the entrance slit) required to produce a given detector response is known. Solutions for other cases can easily be deduced.

Consider a source which radiates in unit time an amount of energy ϵ_λ as a line of wave-length λ, and a continuum whose energy in the range λ to $\lambda + d\lambda$ is $E_\lambda d\lambda$. Let the widths of the entrance, middle and exit slits of the monochromator be respectively s_1, s_2 and s_3; they are assumed to be sufficiently wide for diffraction effects to be negligible. Then, as the magnification produced by the monochromator at each slit is unity, the energy of line of wave-length λ emerging per unit time from the exit slit when the monochromator is set for this wave-length is $\kappa_\lambda s \epsilon_\lambda$, where s is the smallest of s_1, s_2 and s_3, and κ_λ is a factor depending on the aperture of the monochromator and the absorption of its optical system. If the slits are sufficiently narrow for E_λ to be sensibly constant over the range of wave-lengths embraced by them (which should of course be the case) the energy of continuum emerging can be expressed as $\kappa_\lambda E_\lambda . f(s_1, s_2, s_3, \phi_\lambda)$, where f is some function of s_1, s_2, s_3 and ϕ_λ; ϕ_λ is the reciprocal of the linear dispersion of either half of the monochromator at wave-length λ. The form of this function is given in § 3.

The indication of the detector when the monochromator is set at wave-length λ is of course produced by the line and continuum together. Let E_λ' (obtained from the calibration curve) be the energy of continuum per unit of wave-length which would correspond to this detector reading. Then the part of E_λ' to which the line gives rise is $E_\lambda' - E_\lambda$, where E_λ is the part of E_λ' to which the continuum gives rise, which can be obtained by interpolation between continuum values on either side of the line. Since a given detector indication at a given wave-length always corresponds to the same amount of energy incident on the detector, the energy from the line which strikes the detector is

$$\kappa_\lambda (E_\lambda' - E_\lambda) . f(s_1, s_2, s_3, \phi_\lambda).$$

Hence
$$\kappa_\lambda s \, \epsilon_\lambda = \kappa_\lambda (E_\lambda' - E_\lambda) . f(s_1, s_2, s_3, \phi_\lambda),$$

i.e.
$$\epsilon_\lambda = (E_\lambda' - E_\lambda) . f(s_1, s_2, s_3, \phi_\lambda)/s,$$

and thus the energy of the line relative to that of the continuum in the emitted light may be obtained.

In this treatment it has been assumed that there are no lines whose wave-lengths are sufficiently close to λ for them to pass the exit slit (in whole or part) when the monochromator is set at wave-length λ. Measurements can, however, still be made even if this assumption is not justified. It is only necessary to calculate the wave-length setting of the monochromator for the part of the line which is not overlapped, make the observation in this position, and apply the formula given above with the width of the part of the line observed substituted for s. The accuracy of this procedure is, of course, limited by the mechanism of the monochromator.

§3. ANALYTIC DERIVATION OF THE FORM OF THE FUNCTION

Using a geometrical method, Cittert (1926) has obtained expressions for f, but his treatment of the range of validity of the solution is not entirely satisfactory, and the dependence on the dispersion is not mentioned. These criticisms will be amplified after the analytical treatment has been given.

It is first necessary to obtain an expression for the intensity of radiation passing the second slit. There are obviously three cases to consider: $s_1 \gtrless s_2$. Since the width of the image of any wave-length in the plane of the middle slit is s_1 in the first case $(s_1 > s_2)$, the energy of the radiation in the wave-length range dx surrounding any wave-length x in the range $\lambda - \dfrac{s_1 - s_2}{2} \phi_\lambda$ to $\lambda + \dfrac{s_1 - s_2}{2} \phi_\lambda$ will be uniform (namely $k_\lambda E_\lambda s_2 \, dx$), but the energy in the band surrounding wave-lengths outside this range will steadily fall until it reaches zero at wave-lengths $\lambda \pm \dfrac{s_1 + s_2}{2} \phi_\lambda$. Thus the intensity in the range dx surrounding any wave-length x in the regions of varying intensity is

$$k_\lambda E_\lambda \left[\frac{s_1 + s_2}{2} - \frac{|\lambda - x|}{\phi_\lambda} \right] dx,$$

where k_λ is analogous to κ_λ. Hence in the case $s_1 > s_2$ the intensity passing the middle slit is

$$\int_{\lambda - \frac{s_1 - s_2}{2}\phi_\lambda}^{\lambda + \frac{s_1 - s_2}{2}\phi_\lambda} k_\lambda E_\lambda s_2 \, dx + 2 \int_{\lambda - \frac{s_1 + s_2}{2}\phi_\lambda}^{\lambda - \frac{s_1 - s_2}{2}\phi_\lambda} k_\lambda E \left[\frac{s_1 + s_2}{2} - \frac{\lambda - x}{\phi} \right] dx,$$

which comes to $k_\lambda \phi_\lambda E_\lambda s_1 s_2$.

The expression for the intensity in the case $s_1 < s_2$ differs only by interchange of s_1 and s_2, and hence is also equal to $k_\lambda \phi_\lambda E_\lambda s_1 s_2$. This expression holds, again, when $s_1 = s_2$.

The effect of the third slit must now be considered. There are obviously three physically distinct cases:

1. None of the radiation is cut off at the third slit.

2. The third slit only passes wave-lengths which arrive in its plane with the same intensity.

3. The third slit cuts off in the region of diminishing intensity. The conditions satisfied by the slit-widths for each of these three cases can easily be written down since the width of the image of any wave-length in the plane of the third slit is the width of the part of the image of this wave-length which passes the second slit. Using the same principle, expressions for the intensity emerging in each of the three cases can also be obtained.

Thus for $s_1 \gtrless s_2$ the condition for the first case is $s_3 \geqslant 2s_2 + s_1$; the intensity emerging is obviously $\kappa_\lambda E_\lambda \phi_\lambda s_1 s_2$, i.e. $f \equiv \phi_\lambda s_1 s_2$.

For the second case, if (a) $s_1 > s_2$ the condition is $s_3 \leqslant s_1 - 2s_2$; this is again effectively a two-slit case (s_2 and s_3), and so for $s_2 \gtrless s_3$ the intensity emerging is $\kappa_\lambda E_\lambda \phi_\lambda s_2 s_3$, i.e. $f \equiv \phi_\lambda s_2 s_3$. The condition for the second case if (b) $s_1 < s_2$ is $s_3 \leqslant 2s_2 - 3s_1$. This is effectively another two-slit case (s_1 and s_3), the dispersion being twice that of the previous one (since the slits in this case are separated by a double optical system). Hence the intensity emerging for $s_1 \gtrless s_3$ is $\frac{1}{2}\kappa_\lambda E_\lambda \phi_\lambda s_1 s_3$, i.e. $f \equiv \frac{1}{2}\phi_\lambda s_1 s_3$. This case does not occur if $\frac{2}{3}s_2 \leqslant s_1 \leqslant 2s_2$.

For the third case, if (a) $s_1 > s_2$ the condition is $2s_2 + s_1 > s_3 > s_1 - 2s_2$; the wave-lengths at which the slit cuts off completely are $\lambda \pm \frac{s_1 + s_3}{4} \phi_\lambda$, and the wave-lengths between which there is no cutting off are $\lambda \pm \frac{s_3 - s_2}{2} \phi_\lambda$. Hence, by reasoning analogous to that given in the two-slit case, the expression for the emerging intensity is

$$\kappa_\lambda E_\lambda \left[2 \int_{\lambda - \frac{s_1 + s_2}{4}\phi_\lambda}^{\lambda - \frac{s_2 - s_3}{2}\phi_\lambda} \left(\frac{s_1 + s_3}{2} - \frac{2(\lambda - x)}{\phi_\lambda} \right) dx \right.$$
$$\left. + 2\int_{\lambda - \frac{s_2 - s_3}{2}\phi_\lambda}^{\lambda - \frac{s_1 - s_3}{2}\phi_\lambda} \left(\frac{s_1 + s_3}{2} - \frac{\lambda - x}{\phi_\lambda} \right) dx + \int_{\lambda - \frac{s_1 - s_3}{2}\phi_\lambda}^{\lambda + \frac{s_1 - s_3}{2}\phi_\lambda} s_2\, dx \right],$$

which comes to

$$\kappa_\lambda E_\lambda \phi_\lambda \left[\frac{s_2}{2}(s_1 + s_3 - s_2) - \frac{(s_1 - s_3)^2}{8} \right],$$

i.e.

$$f \equiv \phi_\lambda \left[\frac{s_2}{2}(s_1 + s_3 - s_2) - \frac{(s_1 - s_3)^2}{8} \right].$$

If (b) $s_1 < s_2$, the condition is $2s_2 + s_1 > s_3 > 2s_2 - 3s_1$, and the expression for the intensity is

$$\kappa_\lambda E_\lambda \left[2 \int_{\lambda - \frac{s_1 + s_2}{4}\phi_\lambda}^{\lambda - \frac{s_2 - s_3}{2}\phi_\lambda} \left(\frac{s_1 + s_3}{2} - \frac{2(\lambda - x)}{\phi_\lambda} \right) dx \right.$$
$$\left. + 2\int_{\lambda - \frac{s_2 - s_1}{2}\phi_\lambda}^{\lambda - \frac{s_2 - s_1}{2}\phi_\lambda} \left(\frac{s_1 + s_2}{2} - \frac{\lambda - x}{\phi} \right) dx + \int_{\lambda - \frac{s_2 - s_1}{2}\phi_\lambda}^{\lambda + \frac{s_2 - s_1}{2}\phi_\lambda} s_1\, dx \right],$$

which again comes to

$$\kappa_\lambda E_\lambda \phi_\lambda \left[\frac{s_2}{2}(s_1 + s_3 - s_2) - \frac{(s_1 - s_3)^2}{8} \right].$$

The conditions for, and intensity emerging in, this case if (c) $s_1 = s_2$ are given by substituting this relationship in either of the above conditions and expressions, respectively.

To summarize, the conditions for the various cases are therefore:

1      $s_1 \lessgtr s_2, \quad s_3 \geqslant 2s_2 + s_1.$

2 (a) $s_2 \lessgtr s_3, \quad s_3 \leqslant s_1 - 2s_2;$

  (b) $s_1 \lessgtr s_3, \quad s_3 \leqslant 2s_2 - 3s_1.$

3 (a) $2s_2 + s_1 > s_3 > s_1 - 2s_2;$

  (b) $2s_2 + s_1 > s_3 > 2s_2 - 3s_1;$

  (c) $3s_2 = 3s_1 > s_3.$

In his treatment of the problem, Cittert (1926) distinguishes four cases (according to the intensity emerging) and gives the conditions for them as:

A $2s_2 > s_1 + s_3.$

B $2s_2 < s_3 + s_1$ and $s_3 > s_1 + 2s_2.$

C $2s_2 < s_3 + s_1$ and $s_3 < s_1 + 2s_2$ and $s_1 < s_3 + 2s_2.$

D $2s_2 < s_1 + s_3$ and $s_1 > s_3 + 2s_2.$

There are several objections to this. Thus it is not pointed out that case B (corresponding to 1) is significant for $s_1 \lessgtr s_2$; the second condition of case B implies the first. A similar objection applies to D (which corresponds to 2(a)); moreover, it is not stated that D is significant for $s_2 \lessgtr s_3$; that A (corresponding to 2(b)) is significant for $s_1 \lessgtr s_3$ is not mentioned either—the condition given is inaccurate. The statement of the conditions for case C is misleading: if $s_1 > s_2$ (corresponding to 3(a)), the first, which is inaccurate, is irrelevant, and if $s_1 < s_2$ (corresponding to 3(b)), then the last one is irrelevant; mention of 3(c) is omitted. This method of presentation, moreover, somewhat obscures the physical significance of the conditions.

ACKNOWLEDGMENTS

REFERENCES

PERRY, 1938. *Proc. Phys. Soc.*, 50, 265.
VAN CITTERT, 1926. *Z. Instrumkde*, 46, 557.

THE TRANSIENT FLOW OF HEAT THROUGH
A TWO-LAYER WALL

By MIRIAM V. GRIFFITH and G. K. HORTON,*

B.E.A.I.R.A., Greenford, Middlesex

* Now at Imperial College of Science and Technology

ABSTRACT. The sudden introduction of a high temperature source of heat into a room is idealized by supposing that the walls absorb a constant uniform flux of heat. Formulae for the temperature distribution after various times are derived on the hypothesis that a wall is of infinite extent and comprises a uniform finite layer superimposed on an indefinitely thick base. The factors which determine the rate of rise of the wall-surface temperature are particularly studied.

§ 1. INTRODUCTION

THE report gives a mathematical treatment of the problem of the temperature distribution in a non-homogeneous wall of two components under transient conditions when the flow of heat into the surface of the wall is constant. This is of particular interest when considering the rise of temperature of insulated wall surfaces in a domestic room, heated by radiation from a high-temperature source. The particular application envisaged is when such a room is adapted, by means of suitable wall coverings, to promote conditions of comfort in as short a time as possible after applying the source of heat.

From an economic point of view, it is desirable to know the thickness of thermal insulation required, and the effect on the temperature distribution of the thermal properties of the insulation material. By means of certain assumptions, which have been justified experimentally, it is possible to form an equation from which curves can be constructed for practical use.

§ 2. MATHEMATICAL FORMULAE

Assume first that the area of the wall is great, so that the heat is conducted perpendicularly to its surface, which is taken as the plane $x = 0$. The wall consists of two components, one of thickness α, extending from the plane $x = 0$ to $x = \alpha$, while the second component extends from $x = \alpha$ to $x = \infty$ (under practical conditions, and for the times and materials under consideration, this conception introduces no error).

The temperature $U(x, t)$ satisfies the following conditions :

$$\left. \frac{dU}{dt} = D \frac{d^2U}{dx^2} \begin{array}{l} t>0, \quad \alpha \geqslant x \geqslant 0; \quad D = D_1 \\ t>0, \quad \infty \geqslant x \geqslant \alpha; \quad D = D_2 \end{array} \right\} \quad \cdots\cdots (1)$$

$$\underset{t=0}{\mathrm{Lt}}\, U_1(x, t) = 0, \qquad\qquad \cdots\cdots (2)$$

$$\underset{x=\alpha}{\mathrm{Lt}}\, U_1(x, t) = \underset{x=\alpha}{\mathrm{Lt}}\, U_2(x, t), \qquad\qquad \cdots\cdots (3)$$

$$\underset{x=\alpha}{Lt}\, k_1\frac{dU_1}{dx}=\underset{x=\alpha}{Lt}\, k_2\frac{dU_2}{dx}, \qquad \ldots\ldots(4)$$

$$\underset{x=0}{Lt}\left(-k_1\frac{dU_1}{dx}\right)=H \text{ for } 0<t<\infty, \qquad \ldots\ldots(5)$$

$$\underset{x=\infty}{Lt}\, U_2(x,t)=0 \text{ for } 0<t<\infty, \qquad \ldots\ldots(6)$$

where subscript 1 refers to top layer 1, subscript 2 refers to base layer 2, and

$x=$ distance perpendicular to the face of the wall,

$t=$ time,

$D=$ thermal diffusivity $=k/S$.

$k=$ thermal conductivity,

$H=$ constant flux of heat perpendicular to the wall surface,

$S=$ volume specific heat (specific heat $\times$ density).

Let

$$\overline{U}=\int_0^\infty Ue^{-pt}\,dt, \text{ the Laplace transform of } U;$$

then

$$\frac{d\overline{U}}{dt}=p\overline{U}-U(0);$$

hence

$$\frac{d^2\overline{U}}{dx^2}-\frac{p}{D}\,\overline{U}=0 \text{ since } U_2(0)=0. \qquad \ldots\ldots(7)$$

Thus

$$\overline{U}_1=A_1e^{-\sqrt{p/D_1}\,x}+B_1e^{\sqrt{p/D_1}\,x}, \qquad \ldots\ldots(8)$$

$$\overline{U}_2=A_2e^{-\sqrt{p/D_2}\,x}. \qquad \ldots\ldots(9)$$

The boundary conditions give

$$-A_1\sqrt{p/D_1}+B_1\sqrt{p/D_1}=-H/k_1p \text{ at } x=0; \qquad \ldots\ldots(10)$$

also

$$\frac{k_1}{k_2}=\frac{(-A_2/\sqrt{D_2})e^{-\alpha\sqrt{p/D_2}}+(B_2/\sqrt{D_2})e^{\alpha\sqrt{p/D_2}}}{(-A_1/\sqrt{D_1})e^{-\alpha\sqrt{p/D_1}}+(B_1/\sqrt{D_1})e^{\alpha\sqrt{p/D_1}}} \qquad \ldots\ldots(11)$$

and

$$A_1e^{-\alpha\sqrt{p/D_1}}+B_1e^{\alpha\sqrt{p/D_1}}=A_2e^{-\alpha\sqrt{p/D_2}}+B_2e^{\alpha\sqrt{p/D_2}}. \qquad \ldots\ldots(12)$$

These lead to

$$B_1=-\frac{\dfrac{H\sqrt{D_1}}{k_1p\sqrt{p}}\left[e^{-\alpha\sqrt{p/D_2}}-\dfrac{k_2}{k_1}\sqrt{\dfrac{D_1}{D_2}}\,e^{-\alpha\sqrt{p/D_1}}\right]}{(e^{-\alpha\sqrt{p/D_1}}-e^{\alpha\sqrt{p/D_1}})-\dfrac{k_2}{k_1}\sqrt{\dfrac{D_1}{D_2}}(e^{-\alpha\sqrt{p/D_1}}+e^{\alpha\sqrt{p/D_1}})}$$

$$=\frac{-H\sqrt{D_1}}{\gamma k_1p\sqrt{p}}\frac{1}{e^{2\alpha\sqrt{p/D_1}}}\left(\frac{1}{1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}}\right), \qquad \ldots\ldots(13)$$

where

$$\gamma=\frac{k_2S_2+k_1S_1}{k_2S_2-k_1S_1}.$$

Similarly

$$A_2 = (2H/p)\sqrt{D_1 D_2/p}\ \exp.\ \alpha\sqrt{p}(D_1^{-\frac{1}{2}}+D_2^{-\frac{1}{2}})\{\lambda/(1+\gamma e^{2\alpha\sqrt{p/D_1}})\}, \quad\ldots\ldots(14)$$

where
$$\lambda = (k_2\sqrt{D_1}-k_1\sqrt{D_2})^{-1}.$$

Hence

$$\bar{U}_1 = \frac{H\sqrt{D_1}}{k_1 p\sqrt{p}}\left[1 - \frac{1}{\gamma e^{2\alpha\sqrt{p/D_1}}}\left(\frac{1}{1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}}\right)\right]e^{-x\sqrt{p/D_1}}$$

$$-\frac{H\sqrt{D_1}}{k_1 p\sqrt{p}}\frac{1}{e^{2\alpha\sqrt{p/D_1}}}\left(\frac{1}{1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}}\right)e^{x\sqrt{p/D_1}} \qquad \ldots\ldots(15)$$

and

$$\bar{U}_2 = \frac{2H\sqrt{D_1 D_2}\lambda}{\gamma p\sqrt{p}}\left(\frac{1}{1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}}\right)e^{\sqrt{p}\left[\alpha\left(\frac{1}{\sqrt{D_1}}+\frac{1}{\sqrt{D_2}}\right)-\frac{x}{\sqrt{D_2}}\right]}.$$

$$\ldots\ldots(16)$$

Now $\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}$ is essentially less than unity, so $(1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}})^{-1}$ can be expanded as a power series. If so, then, as functions of p, $\bar{U}_1$ and $\bar{U}_2$ will be the sum of terms of the form $(A/p\sqrt{p})e^{-B\sqrt{p}}$. The inverse transforms of such functions of p are known, so that when the expansions have been derived, the inverse transforms U_1 and U_2 may be written down as

$$U_1 = \frac{H}{k_1}\left\{\left[2\sqrt{\frac{D_1 t}{\pi}}\ e^{-x^2/4D_1 t} - x\left(1-\mathrm{erf}\frac{x}{2\sqrt{D_1 t}}\right)\right]\right.$$

$$-\frac{1}{\gamma}\sum_{n=0}^{n=\infty}\left(-\frac{1}{\gamma}\right)^n\left[2\sqrt{\frac{D_1 t}{\pi}}\left(e^{-\{x+2\alpha(n+1)\}^2/4D_1 t} + e^{-\{x-2\alpha(n+1)\}^2/4D_1 t}\right.\right.$$

$$-\{x+2\alpha(n+1)\}\left\{1-\mathrm{erf}\left(\frac{x+2\alpha(n+1)}{2\sqrt{D_1 t}}\right)\right\}$$

$$\left.\left.\left.+\{x-2\alpha(n+1)\}\left\{1-\mathrm{erf}\left(-\frac{x-2\alpha(n+1)}{2\sqrt{D_1 t}}\right)\right\}\right]\right\}; \qquad \ldots\ldots(17)$$

$$U_2 = \frac{2H\lambda\sqrt{D_1}}{\gamma}\sum_{n=0}^{n=\infty}\left(-\frac{1}{\gamma}\right)^n\left\{2\sqrt{\frac{D_2 t}{\pi}}\ e^{-\{x-\alpha\{1-\sqrt{D_2/D_1}(2n+1)\}\}^2/4D_2 t}\right.$$

$$-[x-\alpha\{1-\sqrt{D_2/D_1}(2n+1)\}]\left(1-\mathrm{erf}\frac{x-\alpha\{1-\sqrt{D_2/D_1}(2n+1)\}}{2\sqrt{D_2 t}}\right)\bigg\}.$$

$$\ldots\ldots(18)$$

U_1 gives the temperature at any point x in layer 1 of thickness α at any time t from the introduction of the source of heat, while U_2 gives the temperature in the semi-infinite layer 2 of the wall.

The surface temperature U_0 is given by putting $x=0$ in the expression for U_1, viz.,

$$U_0 = \frac{H}{k_1}\left[2\sqrt{\frac{D_1 t}{\pi}} - \frac{4}{\gamma}\sum_{n=0}^{n=\infty}\left(-\frac{1}{\gamma}\right)^n\left\{\sqrt{\frac{D_1 t}{\pi}}\ e^{-\alpha^2\frac{(n+1)^2}{t D_1}}\right.\right.$$

$$\left.\left.-\alpha(n+1)\left(1-\mathrm{erf}\frac{\alpha(n+1)}{\sqrt{D_1 t}}\right)\right\}\right]. \qquad \ldots\ldots(19)$$

The temperature at the interface, U_S, is given by U_1 when $x = \alpha$:

$$U_S = \frac{H}{k_1} \sum_{n=0}^{n=\infty} \left(-\frac{1}{\gamma}\right)^n \left(1 - \frac{1}{\gamma}\right) \left[2\sqrt{\frac{D_1 t}{\pi}} \, e^{-\left(\frac{\alpha(2n+1)}{2\sqrt{D_1 t}}\right)^2} \right.$$
$$\left. - \alpha(2n+1)\left(1 - \operatorname{erf} \frac{\alpha(2n+1)}{2\sqrt{D_1 t}}\right)\right]. \quad \ldots\ldots(20)$$

§3. APPLICATION OF RESULTS

Representative values of $k_1 S_1$, $k_2 S_2$ for a practical range of materials have been used to calculate the variation of temperature rise of the insulation surface for various times, with the results illustrated in figure 1. This shows the behaviour
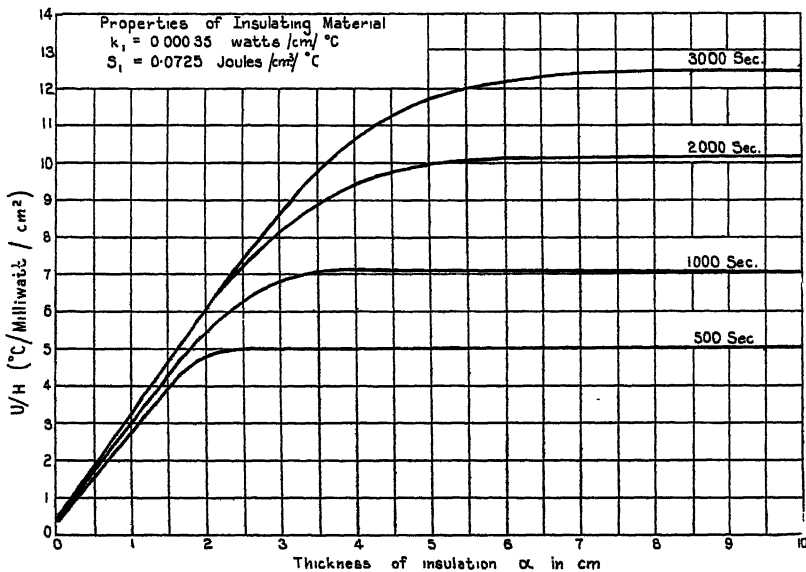
Figure 1.   Variation of surface temperature with thickness of insulation
material on brick.

of the best insulating materials, both as regards high thermal resistivity and low volume specific heat, for which data are available, when used as a wall covering on brick of 5% water content.

It will be observed that for each time of heating there exists a thickness after which the wall behaves as if it were of infinite thickness, and the surface temperature rise corresponding to this is, therefore, the maximum obtainable with that material for the time of heating concerned. Provided this thickness is attained, the base layer 2 may be ignored and the surface temperature will follow the law for a homogeneous wall, viz.,

$$U_0 = 2H\sqrt{(t/S_1 k_1 \pi)} = (2H/k_1)\sqrt{(D_1 t/\pi)}. \quad \ldots\ldots(21)$$

Now in figure 2 has been plotted the thickness α for which the lower layer affects the surface temperature by 10%, and thicknesses exceeding this may be taken as effectively infinite in the application envisaged. This critical thickness is given as a series of curves against $D_1 t$ for various values of $k_1 S_1$, assuming

the lower layer to be brick. For thicknesses greater than the critical value of α, equation (21) applies, so that if a certain value of U_0/H is assumed, the scale of $D_1 t$ can be transformed to a scale of k_1, which has been inserted for a value of U_0/H of $1°$c. per milliwatt per cm.[2] This means that, if it is desired to know the time to reach a temperature rise of $1°$c. per milliwatt per cm.[2], then by noting the value of the thermal conductivity of the wall covering on the scale of k_1, $D_1 t$ is found. If the volume specific heat is known, then D_1 is known and t, the time required, is obtained, while by reading α on the curve corresponding to the appropriate
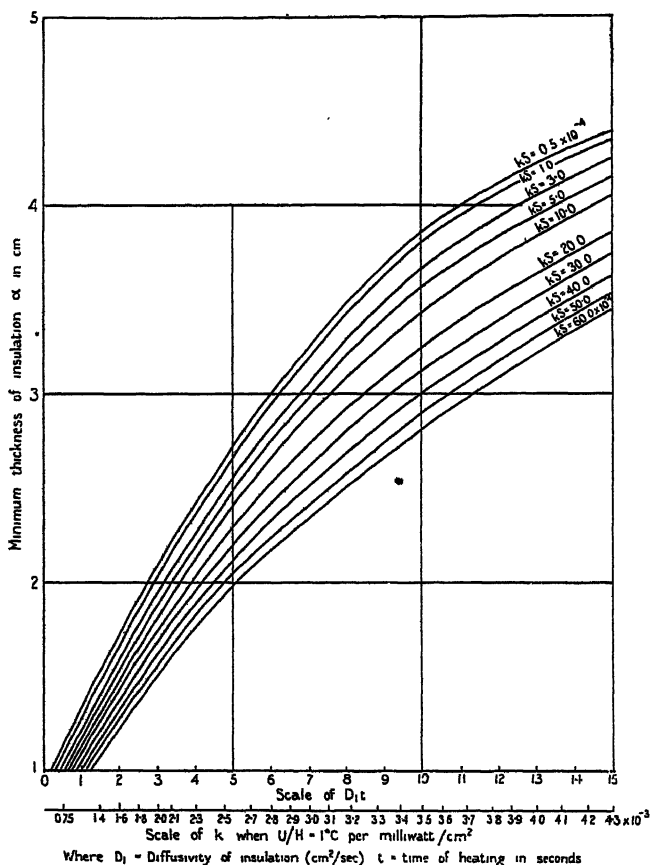
Figure 2.   Variation with Dt and kS of minimum thickness of insulation on a brick wall to give maximum surface temperatures.

value of $k_1 S_1$, the minimum thickness of wall covering can be found. For other values of U_0/H, figure 3 can be used, which gives U_0/H as a function of the time t from equation (21) for different values of $k_1 S_1$. If the properties of the wall covering are known, the time of heating, temperature rise and intensity of the source of heat are given by the curve with the appropriate value of $k_1 S_1$, while figure 2 gives the minimum thickness of covering.

Conversely, if the desired temperature rise and intensity of source (giving the intensity of radiation on the wall surface) are known, then the time to attain this temperature can be found from figure 3 for a series of values of $k_1 S_1$. If any

particular material is chosen to satisfy a chosen value of $k_1 S_1$, then D_1 is also known together with $D_1 t$, which, from figure 2, will then give the minimum thickness, or else, the thickness having been decided, $D_1 t$ can be found from figure 2 for any value of $k_1 S_1$, which again gives the thermal conductivity and volume specific heat of the material required.

As is well known, the thermally insulating properties of a material for transient heating of this type depend ideally only on the product of its thermal conductivity and volume specific heat, decreasing as the latter increases, but the thickness which is required so that, for the given time of heating, the base material does not

Figure 3. Variation of ratio of temperature rise of wall surface to incident radiation with time for a range of values of $k S$ when thickness exceeds minimum (α).

k=thermal conductivity in watts/cm./°c.
S=specific heat in watts/cm.^{3}/°c.

matter, increases with the product of the diffusivity and the time involved. Thus a low value of thermal conductivity or high resistivity is always advantageous, but a low volume specific heat may involve some cost in the thickness of insulation. However, the compromise must take into account considerations other than scientific, so that this aspect cannot further be pursued here.

§4. CONCLUSIONS

(1) With a constant flux of heat a two-layer wall will behave effectively as a single layer, provided the time of heating is short enough in relation to the thickness of the surface layer and its properties. Formulae are derived for the transient temperature distributions.

(2) This critical thickness varies mainly by .increasing with increase of the product of diffusivity of the surface layer and time, for a given base material, and decreases to a less extent with increase of the product of thermal conductivity and volume specific heat,

(3) While for a homogeneous wall the thermally insulating properties in transient heating require a low value of the product of volume specific heat and thermal conductivity, yet a low volume specific heat may increase the thickness of the material required to obviate the effect of a base material of conducting properties on which the first material may be applied.

§ 5. ACKNOWLEDGMENTS

The authors are indebted to the Director of the British Electrical and Allied Industries Research Association for permission to publish this paper, and for his helpful criticism and advice.

DISCUSSION

on paper by R. F. S. HEARMON entitled "The fundamental frequency of vibration of rectangular wood and plywood plates" (*Proc. Phys. Soc.*, **58**, 78, (1946)).

Dr. KANTOROWICZ. It is stated in § 3 that " ... The most powerful response in all cases occurred when the frequency of the driving current was one-half the true resonant frequency ... " This is not surprising, for with the arrangement shown on page 83 the armature A is attracted twice per cycle by the magnet B. There is, therefore, strict resonance between the vibration exciting force and the vibrating body.

This frequency doubling can be avoided by polarizing the magnet either by means of a permanent magnet core or by means of a polarizing D.C.

It should be realized that such polarization is a powerful means for magnifying the alternating force between magnet and armature.

The force F of attraction is proportional to the square of the flux, which in turn is proportional to the sum of the polarizing and the alternating currents i and δi respectively; therefore the alternating rorce is $F + \delta F = a(i + \delta i)^2$.

If the static force, $F = ai^2$, is subtracted it is seen that $\delta F = a(2i\delta i + \delta i^2)$.

Even if δi follows a sine law the force of attraction will contain a term in $\sin^2$. Consequently it is not a pure harmonic force, but contains higher harmonics.

This is the simple explanation of the observation that the plates "respond to integral sub-multiples of the resonant frequency". In short, those sub-multiples have harmonics at the resonant frequency.

It is noted that the author feeds the voltage developed across the magnet to an oscilloscope. It is recommended that instead he should deflect the electron beam by a signal in phase with the current and, therefore, in phase with the exciting force, for in this case he can easily observe whether the velocity of plate motion is in phase with the exciting force. (His pick-up, $C \times D$, is velocity-sensitive.)

This condition of exciting force and velocity of vibration being in phase is a very sharp criterion of resonance and much more easily recognized than maximum amplitude.

With improved technique I expect the author will be able to supplement the information of resonant frequency by the probably even more important one of the damping capacities.

AUTHOR's reply. I am grateful to Dr. Kantorowicz for his comments, and hope to try out the practical suggestions in any future work on vibration problems. The arrangement he proposes for avoiding frequency doubling and for magnifying the alternating driving force has not, to my knowledge, been used in experiments on the vibration of plates, but I have used a similar arrangement, due to Ide (1935) (*Rev. Sci. Instrum.*, 6, 296), to excite longitudinal vibrations in wooden specimens. The most powerful response then occurred with the frequency of the driving current equal to the resonant frequency, in agreement with Dr. Kantorowicz's prediction. The existence of the frequency doubling was, of course, recognized and allowed for in my paper, and the results are, therefore, not affected by it.

REVIEWS OF BOOKS

Mathematical Theory of Elasticity, by I. S. Sokolnikoff. Pp. 394. (Brown University: Summer Session for Advanced Instruction and Research in Mechanics, June 23rd–September 13th, 1941.)

The first reaction to this course of lecture notes must be a feeling of admiration for the American student, the second is likely to be one of pity. Previously equipped with a good working knowledge of applied mathematics, the summer student (and how warm is it in Providence in summer ?) is expected to imbibe six large doses of an elixir compounded of the essences of Love and Timoshenko and laced with rarer precious oils from far-off Tartary. All students who may stay the course will certainly be finished elasticians ; some others may be tempted to insert the conjunction " as ".

Assuming that the title-page is to be disregarded and that the draught is not to be swallowed in such indecent haste, it seems odd that so important a work as this should be given no more than cyclostyled circulation. Even in the course of review the 394-pages have become hopelessly dog-eared, and the reviewer's present preoccupation is to find some way to preserve the tattered remnants in some sort of usable form. For, despite its production and title, this presentation of the mathematical theory of elasticity takes its proper place as a standard work of reference, and the very lowest appraisal must rank it as a valuable supplement to Love and Timoshenko.

Following the King of Hearts' advice, the work starts at the beginning, and the first page clearly states those limitations of the scope of the theory which are too often overlooked. The remainder of the King of Hearts' advice is less well followed, and the scope of the course itself is nowhere clearly delimited. The first three chapters, representing about one-quarter of the whole work, are devoted to analysis of stress, analysis of strain, and the stress-strain relation respectively. These chapters follow more or less classical lines, even to discussion of stress and strain quadrics, but the approach is almost entirely mathematical, and tensor notation is freely used. It may be remarked here that the mathematical background is heavily emphasized throughout, so that the title might well have been the " theory of elasticity for mathematicians " ; the mere engineer may perhaps feel that affine transformation, tensor fields, etc., scarcely pay their passage in the later chapters of application. Some few passages are likely to please neither mathematician nor engineer. For instance, the passage " the fact that the strains components e_{ij} cannot be prescribed arbitrarily can be seen from the following geometrical considerations ", leading after four pages of mathematics to the conditions of compatibility, makes the worst of both worlds. Whilst the passage " It is clear from physical considerations that one must demand that the functions u, v, w (the displacements) be single valued and continuous functions throughout the region occupied by the body " scarcely helps the engineer and cannot possibly satisfy the mathematician. The latter may be mollified later by a separate treatment of multiply connected regions ; but this latter passage is likely only to confuse the engineer, whose shaken confidence may finally disappear entirely in the section on uniqueness of solution.

After deploring the lack of uniformity of notation, the author adopts yet one more : the new notation aims at combining the best elements of all the others, but will uniformity be achieved in this way ? Moreover, in " Young's modulus and the ratio of Poisson " (*sic*), the author strikes an entirely new and surprising discord. A similar perversity is noticeable in the exclusion of aluminium alloys from a list of typical materials ; in the " demand that the resultant force and moment vanish " in one section, followed almost immediately by the dynamical equations ; in the strain energy function inextricably tangled with Hooke's law ; and in the statement that " the *strains* within the body are not susceptible to direct experimental verification ". The author's point would be made equally well by substituting " stresses " for " strains ", so why insult the radiologists ?

The one serious fault in these early chapters is the omission of all but passing reference to large displacements. The last section of Chapter II, on finite deformation, may give

the impression that the two fields are synonymous, particularly because this section opens with the claim that " all the principal results of the classical theory of infinitesimal strain " have previously been derived.

The best features of these chapters are the insistence that crystalline solids cannot be more than statistically isotropic, unless the single crystals themselves are isotropic, and the reserved manner of statement of St. Venant's principle.

Chapter IV starts badly with a discussion of " engineer's theory of bending ", which leads virtually nowhere ; the true evangelist in this field should teach the engineer to remember the shear deflection, not tempt him to forget it. However, having found a firm footing in mathematics by identifying the torsion problem as Neumann's problem, the author canters off (" instead of solving a problem of Neumann, we can equally well solve a problem of Dirichlet ") into a long and very valuable digression on the uses of the complex variable and conformal mapping. The engineer who has to struggle with the mathematician's w- and z-planes may feel a savage desire to retort with a blue-print or two ; but this section is worth while, and the engineer should remember that the author is being as kind as he may be. On the other hand, the engineer may relieve his pent-up feelings on the section on curvilinear co-ordinates. Too high a price can be paid for generality, and this section is bought in the black market ; and surely curvilinear co-ordinates should be related to the complex variable !

At the end of the chapter the author returns to the " engineer's theory of bending ", only to fail spectacularly to retrieve his initial fault. By way of recompense there is an excellent appendix on the Schwarz-Cristoffel transformation (which, incidentally, is one of the few mathematical tools the student is not expected to have already mastered).

Chapter V deals with the problems of plane strain and plane stress, which are shown to be identical. Once more solutions are developed in terms of complex variables, and the treatment appears thorough ; but the going is hard, and the engineer might be encouraged by more remunerative results than examples on circular and elliptic contours. But at least the connection between complex variables and curvilinear co-ordinates is now made moderately clear.

Chapter VI, on the deflection of plates, deals extensively with the aeolotropic plate in deference to the technical importance of plywood and fibre-filled plastics ; in view of other omissions this inclusion seems odd. Moreover, the whole section wears (to the reviewer) an unreal air, because large deflections creep in only towards the end ; and a reference back to Chapter II for the basic formulae is not in the least helpful for the reason already noted. The mode of treatment of the large deflection problem and the example chosen afford but a poor distorted glimpse of this field.

Undoubtedly the chief value of the work as a whole lies in its applications of the complex variable. These methods deserve to be better known, and the present exposition should prove a useful reference. Yet the mathematician does not sell his goods particularly well, and many who could use these methods will not master them from this treatise. If this stricture be judged hard, let the engineer ponder the author's definition of a plate : " A cylinder is called a plate if its height is small compared with the linear dimensions of its cross-section " : and to think that the reviewer has meekly accepted criticism for speaking of the " slightly curved plate " ! H. L. COX.

Infra-red and Raman Spectra of Polyatomic Molecules. by GERHARD HERZBERG, Pp. iv + 632. (New York: D. Van Nostrand Co., 1945). $9.50.

In an important monograph published some years ago, Professor Herzberg gave us a thorough survey of the principles underlying the different kinds of spectrum associated with diatomic molecules, and their relevance to physical and chemical problems. A second similar volume had been planned to deal with the spectra of polyatomic molecules. As might have been expected, however, the rapid accumulation of experimental results, as well as the related theoretical background, made it necessary to divide the volume into two parts. The present book on infra-red and Raman spectra forms the first part, and a further volume is promised which will deal with the ultra-violet spectra.

Any reader of this book, not least the spectroscopic expert, will be deeply impressed by the vast amount of material collected by the author and by the clear way in

which he has set it out. Several other recent monographs have dealt with either the infra-red or the Raman spectra, and many experimental and factual details can be found in them which are not included in the present book. However, the close relationships between the two types of spectrum make it desirable that, for many purposes at least, they should be considered together, and in this book the author has given us an admirable correlation of the data, as well as the most complete record of infra-red measurements yet published.

The infra-red spectra of even the simplest polyatomic molecules are far more complex than those of diatomic molecules. This arises because of the very marked increase in the number of possible molecular energy levels as the number of atoms in the molecule increases, and in their highly complex arrangement as regards magnitude. Elucidation of the spectra demands a detailed understanding of the energy states, and the rules governing spectral transitions between them. If the exact relationships are known between the values of the molecular energy levels on the one hand and the molecular dimensions, nuclear masses and electronic structures on the other, it may be possible to derive important structural information from the spectroscopic data. Even when the quantitative correlations are incomplete, useful qualitative information is often obtained.

The author has discussed all these matters exhaustively and systematically, and although in such a masterly presentation the reader's path must necessarily be hard, the end fully justifies the journey. The general reader may be surprised to find a substantial amount of ancillary apparatus here for use in the theoretical treatment. For example, there is a lucid, though concise, account of symmetry operations and point groups, and a detailed treatment of the mathematical technique for dealing with the harmonic vibrations of molecular models using symmetry or normal co-ordinates. As regards the symmetry properties, the arguments are illustrated by clear diagrams, and examples are given of molecules belonging to the different point groups. From the point of view of the physicist or chemist this is perhaps the most satisfactory and usable summary of this subject yet published. The methods given for calculating the normal vibration frequencies of a molecule on the basis of a given potential energy function illustrate not only the importance of the different types of force field which can be assumed but also the procedure for solving the secular equations which arise, and the simplification of such equations by the introduction of conditions of symmetry. The author has reviewed comprehensively all existing information obtained by applying such calculations to simple molecules whose frequencies of vibration are known. This makes it possible to assess in proper perspective the values obtained for the force constant of particular linkages in different molecules.

These discussions on normal vibrations lead naturally to the consideration of other related matters such as torsional motions and inversion doubling found with pyramidal structures like ammonia. The author also gives an informative account of the product rules of Teller and Redlich which relate the frequencies of a pair of isotopic molecules to each other independently of the particular potential energy function which may apply. In about a hundred closely packed pages full references are given to work on the vibrational spectra of all the important molecules yet studied, and there is also a critical discussion of the significant points which have emerged. This compilation alone is very valuable, but it also serves to bring out some of the special cases which require further examination.

There follows a detailed consideration of interactions between vibration and rotation, illustrated by the data for many vibration-rotation spectra. The important phenomenon of Coriolis interaction is discussed in some detail, although here the specialist may regard some of the interpretations as problematical. The remarkable increase in complexity of the rotational fine structure of vibration bands of asymmetrical molecules compared with linear molecules or symmetrical tops is illustrated both by the theoretical considerations of the energy levels and by the actual examples chosen. The latter serve to emphasize the point that even a small departure from the symmetrical top may lead to markedly increased complexity of the rotational fine structure. Most of the examples discussed in detail deal with resolved rotational fine structure. In reality, however, asymmetrical molecules are the majority, and most of them have moments of inertia which are too big to lead to such a resolved structure of the rotational branches. In such cases band contours are observed, which will be hybrids if the vibration concerned has components of change of electric moment along more than one of the axes of inertia. The question then arises whether these contours can be used as reliable guides in the assignment of vibrations to particular types, or

even in the approximate determination of moments of inertia. The author has summarized briefly the theoretical computations of the appropriate band envelopes to be expected for different degrees of asymmetry, but in view of the growing number of experimental measurements in this connexion a fuller account might have been welcome.

There is a final chapter entitled *Applications*, confined almost entirely to a polished account of the calculation of thermodynamic properties from spectroscopic data. The book concludes with an excellent bibliography and index.

Some may feel, after reading this book, that, admirable as it is, it still remains somewhat cold, rigidly austere and perhaps a little idealized in its treatment, for it deals to a large extent with the derivation of the energy levels of molecules having some degree of symmetry, and also having small moments of inertia and being capable of examination in the vapour state, where the rotational fine structure of the bands can be obtained. This criticism may be valid, and has some point in the sense that the majority of molecules in practice are asymmetrical rotators, have large moments of inertia, and are comparatively involatile. But the author's failure to deal with these molecules in the same exhaustive way is not at all detrimental to the book as it stands. The division is in some respects the division between physics and chemistry. In order to make full use of infra-red and Raman spectroscopy in chemistry, where the bigger and unsymmetrical molecules have the stage, an entirely different and more empirical approach may be required. Much indeed can be learnt in this way about intermolecular forces such as hydrogen bonding, and much can be done by the application of infra-red and Raman measurements to the question of structural diagnosis for special intramolecular groupings. Also the whole new field of infra-red analysis, both qualitative and quantitative, has recently been developed. A proper discussion of all these matters would have seriously lengthened the book, and in any case many of them are in a state of rapid growth and flux.

This book is a landmark for those directly interested in molecular spectroscopy. Others, both physicists and chemists, wishing to understand how infra-red and Raman spectra can be applied to the determination of molecular vibrations and the molecular structure of simple molecules will gain much from it. Both the author and the publishers are to be congratulated upon its production. H. W. T.

An Index of Mathematical Tables, by A. FLETCHER, J. C. P. MILLER and L. ROSENHEAD. Pp. vii + 450. (London: Scientific Computing Service, Ltd., 1946). 75s.

A worker who requires a figure from a mathematical table may desire either to work out how some other function varies, or may need just a particular value, for example the reciprocal of some number which occurs in his own calculations. In the former case, the need for an orderly table is evident ; in the latter, the user needs a table so that he may interpolate for his particular argument. The degree to which a table is necessary (as opposed to convenient) varies greatly. If one wants a reciprocal, it is to-day practically as easy to obtain it from a calculating machine as from a table ; if the value to be calculated is a wave-number from a wave-length, so that the reciprocal has also to be multiplied by a given factor, then the table of reciprocals is really of little value. A table giving the conversion direct is of value, but a table of reciprocals would need to be supplemented by a table of products or else both replaced by a table of logarithms. Again, when a calculation presents a sine as its answer, the argument is likely to be in radians, and numerical calculation from the formula demands either a table of sines with the argument in radians, or a table for converting radians to degrees. In navigation, on the other hand, the angles will nearly always be given in degrees, minutes and seconds, or occasionally in hours, minutes and seconds. One table of sines will serve all purposes if the user is prepared to carry out the conversion of the argument to a standard system, but it is undeniable that a great deal of time is saved in the aggregate if there are different tables for the different purposes.

Against these examples, of tables which are convenient but not indispensable, we may set those where the work really cannot be done if tables have not been prepared. Such cases arise where the needs are for factors of large numbers, or for values of Bessel or gamma functions or their inverses. The flow of heat or electric current in a cylinder

involves Bessel functions, and if an experimentalist desired to calculate the constant of a material cylinder from observations on current flow in it, his task would be impracticable if he had to calculate the necessary functions instead of taking them from tables.

There are functions of an intermediate type, exemplified by $(1-x^2)^{\frac{1}{2}}$. This is not difficult to calculate by machine, especially if $1-x^2$ is factorized. It may also be obtained from tables of squares and square roots, or yet again from tables of sine and cosine (in any argument, provided it is the same for both). Yet many workers have thought it worth while to tabulate the function, and evidently a direct table must save much time if many values have to be worked out. Tallqvist seems to have been the first : he gave the function to 12 decimal places, at intervals of only 0·01, so that a good deal of interpolation would be called for. Fourteen years later, in 1920, T. Smith gave a four-figure table at intervals of 0·001 (0·0001 near unity) ; he was the first to give the table from $x=0·7$ to $1·0$, which enables the reader to obtain all values, by using the table as one of inverse functions over the rest of the range. Since 1920, according to the lists in the volume now under review, six other tables have been computed or published, to various degrees of accuracy. None of them follows Smith in his space-saving device.

Whichever type of table is needed by an investigator, whether the essential type or the convenient, it will be of no use to him unless he can learn quickly of its existence and location. The *Index* now under review (it will, I think, become known to mathematicians as the *Liverpool Index*) sets out to help in this problem. It consists of two main parts, of which the first is a list of functions, with the tables of each function set out in order. For each such table the *Index* sets out the accuracy (number of decimal or significant figures), range and interval of the argument, information about the means of interpolation (whether differences are tabulated and, if so, what orders, and whether proportional parts are shown), and finally the reference, in the form of the author's name and the year. In the second part, all authors are set out in alphabetical order, with their different publications by date, so that the references in the first part become precise by reference to this second part. In addition to the main parts, we should mention an interesting introduction by the two authors whose names stand first, another by Prof. Rosenhead (whose war work detached him from most of the work, after it had been planned), a publisher's preface by Dr. Comrie and a foreword by Prof. Hartree.

We learn in the preface that the authors do not believe they have attained completeness, and even that they could have approached nearer to that ideal if conditions had been different in the last six or seven years. This is no doubt true, but there are no striking omissions ; the reviewer has compared the references in part 2 for the letters A and B with his own notes, and found far more in the *Index* than in his MS. The only converse cases are a publication in 1859 of a table by C. D'Aiguières, *Tables sans fin donnant les résultats de la multiplication, de la division et de l'extraction des racines carrées et cubiques de tous les nombres imaginables* ; J. Blater and A. Steinhauser, 1888, *Multiplication and Division rendered rapid and easy by the calculation table, giving the 9 multiples of all numbers* ; and J. Brown's *Mathematical Tables* (Edinburgh, 1830). None of these can be of any importance. I have no details of the French work ; there are plenty of multiplication tables, and the collection by Brown, although it ran to at least five editions, was not a specially good one.

The need for the volume is really very great, and as far as can be judged without a year or so of trial, it has been extremely well done. That is, the choice of functions deliberately omitted is wise—there is no reason to suppose that there are important accidental omissions—and the information given seems to be accurate. The volume is really well produced, and for this it is clear that the mathematical world owes much to Dr. Comrie, who has given not only time and trouble, but has himself, as Scientific Computing Service, borne a very considerable expense (running to some thousands of pounds). To say that every scientific library needs the book is an understatement. It should, despite its cost, be bought by as many working mathematicians and mathematical physicists as possible.

J. H. A

THE PERFORMANCE OF AIRCRAFT CAMERA LENSES

BY E. W. H. SELWYN AND J. L. TEARLE

Communication H.1072 from the Kodak Research Laboratories, Harrow

ABSTRACT. The estimation of the best combination of lens and film for aerial reconnaissance photography is not a simple problem, owing to the number of variables, such as focal length and aperture of the lens, the size and speed of the film, forward motion and speed of the aircraft which are involved. For quantitative considerations it is necessary to adopt a measure of the detail-revealing capacity of the negatives, and for this purpose the resolving power was chosen, since it was not only obviously related to the capacity of the negative to reveal detail, but also capable of justification as a possible means of estimating the performance of the lens in optical image formation. In order that the resolving power figures may correspond as closely as possible to practical performance, careful choice of test-object was necessary, for experiment showed that different test-objects gave different results. A test-object consisting of two rectangles separated by their width, with a difference in density of 0·2 between them and the background, was adopted. Apparatus with such test-objects at the focus of a collimator of long focal length was set up, and with it measurements were made of the photographic resolving power all over the field, and in a set of planes embracing the focus, for a representative collection of air camera lenses. The results showed that the photographic resolving power improved as the lenses were stopped down from their maximum apertures until very small apertures were reached, and that as a general rule the resolving power diminished towards the outer parts of the field. The variation with aperture was accounted for, in general terms, by theoretical considerations, and certain regularities in the change of resolving power with angular distance from the optical axis were also explained. A figure for the performance of any lens was obtained by averaging the resolving power over the field of the lens, having regard to the greater area covered at the larger angles. It was found that the performance of any of the lenses could be represented with very fair accuracy by the formula

$$R = \left[\frac{207}{fG}\right]^{\frac{1}{4}} \left[\frac{F.\,No.}{\tan^2 \theta}\right]^{0\cdot3},$$

where R is the mean resolving power in lines per mm. over the field up to an angular distance θ from the axis, f is the focal length of the lens in inches, and $F.\,No.$ its relative aperture. G is the granularity of the film in density-microns at a density of 1·0.
This formula has no substantial theoretical basis, although it can be accounted for to some extent. It has proved useful, for it assists in the choice of the focal length, aperture and angle of view most appropriate for any given purpose in aerial reconnaissance photography. It may also be used, at least until the progress of lens design renders it inaccurate in form, for estimating the merit of a given lens design by comparison of its performance with that given by the formula.

§1. INTRODUCTION

IN order to make clear the reason for examining the problem with which this paper deals, a brief sketch is given in this *Introduction* of the work on aerial photography of which it forms part.

Soon after the war began it seemed likely that photography from aircraft would be an important item in it. In spite of the considerable development of aerial photography in the war of 1914–1918, the problem of estimating what were the best possible results, and how they could be secured, still seemed to be largely unsolved, especially on the theoretical side. Consequently, very early in the war, the Kodak Research Laboratories took much interest in this problem in order to discover what sort of film was best for aerial photography. Preliminary discussions soon showed that it was one of some complexity. The main use of aerial photography was likely to be in reconnaissance, in which it is important to record the utmost possible detail. The other use, for map making, requires distortion of the image introduced both by the lens and by the deformation of the film base to be reduced to the minimum, and presumably some loss of definition would be tolerated if that were the necessary price to pay for absence of distortion. Aerial survey for map making, however, did not become an urgent problem till later in the war, and this paper is concerned with one broad aspect of the general problem of securing the best possible definition in aerial photographs.

The nature of the general problems may be gathered from the following considerations. It is necessary to give at least a certain minimum exposure if the negative is to be of value. Thus the time of exposure T, the focal ratio $F.No.$, the brightness of the subject B, and the speed of the film S must be in a very definite relation. This relation, in fact, is

$$S \propto (F.No.)^2/BT.$$

To see that this is so, it is only necessary to note that it constitutes the practical definition of S. For a given value of B, therefore, a wide variation in values of S, $F.No.$ and T is possible, which will, nevertheless, satisfy the condition that the exposure must be adequate. But out of the possible sets of values of S, $F.No.$ and T some will lead to sharper photographs than others, as may be seen in the following way. It is a matter of common experience that sharper photographs are obtained with slow films than with fast ones. But if the speed of the film is decreased then either T must be increased or $F.No.$ decreased, or both. An increase in T, however, necessarily causes an increase in the blurring of the photograph by the forward motion and vibration of the aircraft. Also it is a matter of common experience that a moderate reduction in the aperture of photographic objectives of normal construction improves the sharpness of the photographs taken with them, so that a decrease in $F.No.$ may be expected to diminish the sharpness. The increase in sharpness expected from the use of slow film is therefore offset, more or less, by the increase in time of exposure or aperture, or both. Hence, unless a remarkable exactness of balance occurs between the various effects involved, we must expect that some particular set of values of S, $F.No.$ and T will lead to the sharpest photographs.

The foregoing argument merely demonstrates that there is probably an optimum as regards sharpness. In order to understand how this optimum is determined it is necessary to know how the properties of the film affect the definition and how the optical properties of the lens are affected by a change in aperture.

A little consideration will also show that we must know more about the lens than the variation of its properties with aperture. On general grounds it is clear that

the larger the image on the film can be made, even if the faults in the image given by the lens are magnified in proportion, the less serious will be the relative degradation of the image due to the film. We are thus led to expect that better results might be obtained with lenses of long focal length than with short. For a given size of film, however, the area of ground covered in a single photograph is less the longer the focal length of the lens. From an operational point of view it might be necessary to cover a large area of ground not only for estimating, for example, the damage suffered by a town from aerial bombardment, but also because it is not easy to get a given target into the centre of the picture under operational conditions. Thus it is also important to know whether the angle of view for which a lens is designed has any important effect on the definition.

Two considerable gaps in our knowledge thus existed. One was the way in which the film influences the definition, and the other the way in which the definition obtained with a lens of given design depends upon its aperture, focal length and the size of the film with which it is used which determines its operative angle of view. Until these gaps were filled it could not be asserted with certainty that the best was being achieved with available equipment, nor was it possible to indicate the most profitable directions in which improvement could be made. It might be suggested that this difficulty could be overcome, in empirical fashion, by making aerial photographs with every possible combination of lens and film. It is, however, neither easy nor inexpensive to make reliable tests in an aircraft, and the amount of experiment to arrive at an incontestable conclusion, when so many variables are involved, would be prohibitive.

To begin with, a purely theoretical attack was made. It was hoped that it might be possible so to relate the resolving power of a photographic lens, as observed visually, for preference, and the resolving power of the film, as found with a lens of extremely high resolving power, or with a test-object in contact with the film, that the resolving power of the combination of photographic lens and film could be estimated. The attempt proved abortive. Experiments carried out in the Kodak Research Laboratories, and a further more extensive series carried out by Mr. R. J. Hercock of the Ilford Laboratories, revealed no relation of the sort expected. Mr. Hercock, in fact, did discover an empirical relation which fitted his results, but the value of this for the desired use seemed doubtful. However, further theoretical examination of the problem, on similar lines to the original attempt, but this time bringing in the grain of the film explicitly as a factor disturbing the recognition of detail, suggested a relationship between the resolving power obtained in the negative and the quality of the image formed by the lens, the contrast, turbidity and granularity of the film, and the contrast of the object photographed, which was not in serious disagreement with any of the known facts. Previous to this work no satisfactory reason had been found, for instance, for the fact that photographic materials display a maximum in the curve of resolving power against exposure or density. The turbidity of the emulsion layer had been recognized as a feature reducing the contrast of the image formed in the layer, and it had been recognized that on development this contrast would be increased or decreased according to the slope of the (density : log. exposure) curve of the material, at the point of exposure. Thus the resolving power would be expected to increase with density and then to remain constant throughout the region where the slope of the (density : log. exposure) curve remains constant (the straight-line

region of the characteristic curve). In fact, the resolving-power curve exhibits a maximum near the commencement of the linear region of the characteristic curve. This was accounted for by the continual increase of granularity with density, because an increase in granularity reduces the visibility of any given density difference in the developed image. Advantage was taken of the recognition of this property to construct an instrument for measuring granularity, which proved of great value later on in the work on aerial photography. An account of this instrument has already been published (Romer and Selwyn, 1943).

By means of this theory it was possible to account, qualitatively at least, for the very great differences frequently observed between the resolving powers obtained visually and photographically, for the fact that the resolving power in prints was not greatly affected by the contrast of the printing paper, and for the considerable increase of graininess of prints made from negatives taken under hazy atmospheric conditions compared with those taken under clear conditions. It was realised however, that more exact establishment of this theory would require much time. For the time being, therefore, the theory was used only as a guiding principle, and attention was concentrated on objectives of greater practical importance.

The initial attempts to estimate the effect of varying the aperture, angle of view and focal length were speculative and gave only a fragmentary idea of the nature of the variation. Although, therefore, they suggested the general direction in which the definition might be expected to vary, as the aperture, angle of view and focal length were changed, it was clear that many tests on lenses would be necessary before it could be said that the performance of camera objectives in aerial photography was understood. Some of the early arguments are used, in a form modified by later experience, in the body of this paper.

During the time these attempts at the problem had been in progress, numerous discussions had taken place with the staff of the Royal Aircraft Establishment and some degree of agreement had been reached on the type of work which could be carried out in the laboratory, in contrast with that which must necessarily be done in aircraft. At this stage, the Kodak Research Laboratories were placed under contract to the Ministry of Aircraft Production, with a fairly definite programme of investigations in view. Foremost among these investigations was the establishment of a technique of photographic testing of air camera lenses by which their performance in aerial photography could be estimated. It is with this problem and the results of systematic tests of all the lenses available that this paper is concerned. The results obtained enabled a test to be made of the belief, for which only slight and doubtful evidence was originally available, that the differences between lenses could be accounted for, for the most part, by differences in angle of view, aperture and focal length. In other words, given the focal length, aperture and angle of view, it should be possible to estimate with moderate accuracy the definition given by the lens. It will be seen that this was so for the lenses available for test. This general formula has not the validity of a law of nature ; it is a concise expression of the state to which lens design had arrived by about 1943, and no doubt departures from the general approximate rule which has been found will be more and more frequent in the future as the art of designing lenses progresses. Already some notable advances had been made in this direction.

The general formula connecting the resolving power with the aperture, angle of view, and focal length which was arrived at by this investigation would enable

an estimate to be made of the best lens to be used under any given conditions, provided that the exposure time was sufficiently short to allow the effect of movement of the camera during flight to be neglected.   But as soon as it was seen that the general formula in question was a fair representation of the performance of lenses on the ground, a laboratory investigation was begun on the effect of the forward motion of the aircraft upon the definition of negatives.*   By adopting the general formula arrived at in the survey of lenses (with some limitations), using certain results obtained in subsidiary work concerning the variation of granularity with speed of films, and the information embodied in exposure tables worked out by the Royal Aircraft Establishment, it was found possible to express the definition obtained in aerial photographs as a function of all the " practical " variables entering into the problem, such as speed and height of aircraft, speed of film, aperture of lens (itself determined by the time of exposure and intensity of illumination of the ground), angle of view, and focal length of the lens.   This formula is capable of use in so many different ways, according to the information required, that it has not been subjected yet to exhaustive practical test.   One interesting feature of this particular work, however, was that it suggested that the best time of exposure, for anything approximating to normal conditions, was that already in use by the Royal Air Force.   It is not certain whether this result should be regarded with most satisfaction by the laboratory workers or by those who arrived at the exposure by severely practical considerations.

In addition to the above work, several investigations of interest have also been carried out.   These relate to the spread of light in the emulsion layer caused by the scattering from grain to grain, a check on the theory of resolving power of emulsions for the type of test-object used in the lens tests, and the nature of the correlation between speed and granularity of films for air photography.   An investigation has also been made on the effect of obliquity of the incident light on the resolving power of an emulsion.   Finally, the work on the interaction between the image given by the lens and the film, which leads to the observed resolving power, has been pursued to a state at which we can say with some confidence how the interaction comes about and what determines the resolving power in the negative.

§ 2. RESOLVING POWER AS A MEASURE OF PERFORMANCE

In the early stages of the work doubts were expressed whether measurements of resolving power could be used satisfactorily as a measure of performance. These doubts appear to have arisen, in the main, from acquaintance with the results of resolving-power tests using test-objects of high contrast.   It is known, for instance, that the visual resolving power of a telescope, for objects of high contrast, may be a misleading guide to the state of correction of the lens.   The same criticism can also apply to photographic tests, especially if, in addition to the contrast of the test-object being high, a fine-grain high-contrast photographic material is used.   In such cases also the exposure is very important.   This may be demonstrated by a modification of Jones's tone-reproduction diagram (1921). The curve A in figure 1 (a) represents the intensity distribution in the image of a point source of light, such as might be obtained from an imperfectly corrected lens.   In figure 1 (b) this has been plotted on a logarithmic scale of intensity in

* This investigation is not described in this paper ; it was carried out by Mr. J. M. Gregory, and will be published later.

the lower right-hand quadrant. By projection on to the characteristic curve of the photographic material (shown in the upper right-hand quadrant) the resultant density distribution may be obtained as shown in the upper left-hand quadrant. The full radius of the disk of light produced by the lens is 5 units, whereas under the particular conditions of exposure chosen for illustration, the full radius of the image is only 2 units. Such a possibility is not unknown in practice. Sharpness beyond expectations, based only on the state of correction of a lens, may be obtained on suitable material by careful attention to exposure.

In order to explain how it is that resolving-power tests can be an adequate guide to performance in many cases, and in aerial photography in particular, it is convenient to use the theory of resolving power mentioned in § 1. That theory, to be stated precisely, requires that the test-object be in the particular form of long parallel lines across which the brightness varies as a sine-curve. The main effect of lack of sharpness is to decrease the contrast between light and dark parts of the image, in such a way that the smaller the scale of the image the greater is the reduction of contrast. If the contrast in certain details is above a certain minimum then those details will be seen, but if below, they will not. In more definite terms, if the detail is in the form of the above sine-curve test-object and the contrast is
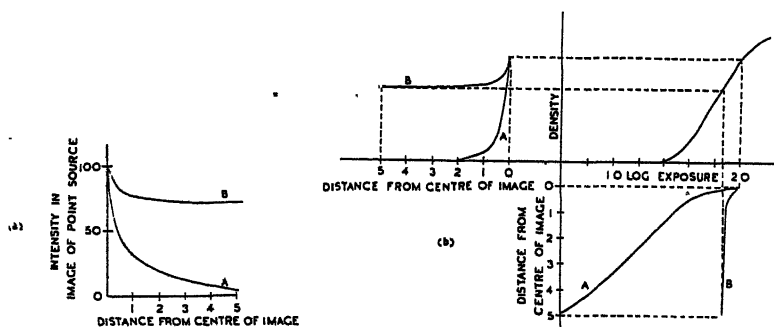
Figure 1.

(a) Schematic representation of intensity across image of point source as given by imperfect lens (Curve A). Curve B represents the intensity distribution when the point source is surrounded by a uniformly illuminated field.

(b) Tone-reproduction diagram showing derivation of density distribution in the developed photographic material.

measured in suitable fashion, the curve of contrast against the logarithm of the reciprocal of the size of detail will follow a course such as that shown in figure 2. The appropriate measure of contrast is $\log P$, where $(1+P)/(1-P)$ is the ratio of maximum to minimum brightness in the image. The value of $\log P$ is approximately equal, for values of $\log P$ up to $\bar{1}\cdot 8$, to $\log \gamma + \log a + \log P_l + \log P_e$, where γ is the slope of the curve of density against log exposure for the photographic material, $(1+a)/(1-a)$ is the ratio of maximum to minimum brightness in the test-object, $\log P_l$ is a function of d (the separation between neighbouring pairs of dark lines in the image of the test-object) and is determined by the image-forming properties of the lens, and $\log P_e$ is a similar function of d which describes how the contrast is reduced by the scattering of light in the emulsion layer of the photographic material.

If the aerial image is being viewed, P_e is unity and γ is unity, there being no

photographic material involved, and the threshold contrast is represented by a straight line parallel to the axis of $\log 1/d$ at a value of $\log P$ of approximately $\bar{2}\cdot3$. The threshold for the photographic case is represented approximately by a straight line at a slope of 3/4,* the ordinate of which at $\log(1/d)=0$ is equal to $\log G^{\frac{3}{4}}-2\cdot0$, where G is the granularity of the photographic material in density microns. The reason for the increase in the threshold contrast for smaller detail is that the granularity becomes relatively more obtrusive and disturbing to recognition of detail the finer the detail.

Where a $\log P$ curve intersects a threshold line, the contrast of the image is equal to the contrast threshold of the receptor (either the eye or the photographic material) and the value of $1/d$ at this point is equal to the resolving power. Thus in figure 2 the intersection of the $\log P_l$ curve with the threshold contrast line for the eye gives the visual resolving power of the lens; the intersection of the $\log\gamma+\log P_e$
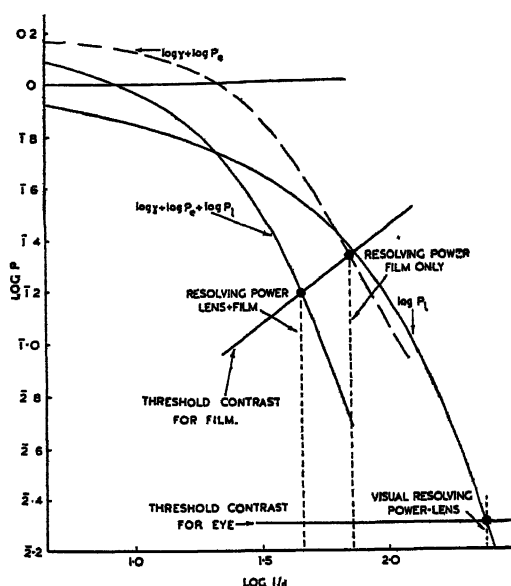
Figure 2. Contrast in visual and photographic images, and thresholds of contrast plotted against reciprocal of size of image (logarithmic scales).

curve with the threshold contrast line for the film gives the film resolving power, and the intersection of the $\log\gamma+\log P_e+\log P_l$ curve with the threshold contrast line for the film gives the resolving power of the lens and film together. For a set of test-objects of different contrast, the $\log P$ curves will consist of a set all of the same shape (namely that of the curve $\log P_l+\log P_e+\log\gamma$), but displaced from each other vertically by the differences in $\log a$ for the test-objects. If, therefore, we are prepared to make resolving-power tests on a number of test-objects of different contrast, we are able to reconstruct the curve of $\log P_l+\log P_e+\log\gamma$. If the observations are visual the reconstructed curve will be $\log P_l$; if the observations are photographic a knowledge of the γ-value and of the light-scattering properties of the photographic material will enable $\log P_l$ to be obtained from

* These figures are derived from those given by Romer and Selwyn (1943) and apply therefore to Cobb test-objects. It is known, however, that approximately the same values apply for sine\} curve or other test-objects.

$\log P_l + \log P_e + \log \gamma$. It is assumed also, of course, that the granularity of the film is known, otherwise the appropriate threshold line cannot be drawn. Log P_l is determined only by the optical properties of the lens; it can be calculated from the light distribution in the image given by the lens of a point source.

Although the method of presentation may be unfamiliar, a set of measurements of resolving power for test-objects of different contrast is equivalent, therefore, to a description of the image-forming properties of the lens, and in that respect is not markedly inferior to any other description. Indeed, it is difficult to see what other measurable index of performance could be used: at any rate, no other has been suggested.

The diagram in figure 2 also indicates how far above the threshold contrast the actual contrast is, and to that extent is a measure of the visibility of objects which are greater in size than at the resolution limit. Now in aerial photography for reconnaissance, the quality of a lens is determined, very largely, by the smallest details visible. If details of a certain size and contrast can be recognized, larger details of the same contrast, or details of the same size and greater contrast, will certainly be recognized, and it is not of much importance whether they can be recognized with a certain facility or whether they can be recognized with a little greater facility. In other words, a resolving-power measurement is likely to be an adequate guide to the performance of an aerial camera lens. Before leaving this discussion it is worth interpolating several conclusions to be drawn from it, not directly connected with the immediate problem of estimating lens performance. It will be seen that it is not possible to derive much information about the photographic resolving power from the visual resolving power for a test-object of given contrast, or vice versa, even if the properties of the photographic material are known in detail. If we know only the resolving power of the film, measured, for instance, by a test-object in contact with the emulsion layer, the situation is even worse. Nevertheless, from a knowledge of the general properties of the curves of $\log P_l$ and $\log P_e$ against $\log(1/d)$ rough estimates or limits of resolving power can be obtained from limited data.

Choice of test-object

So far the discussion of the possibility of using a resolving-power measurement to assess the performance of lenses has been based, where exact quantitative statements are concerned, on the assumption that the test-object is of the sine-curve type. But, in fact, detail of such a character must be rare—the only example approximating to it which comes to mind is a corrugated iron roof upon which the sun is shining not too obliquely. The question therefore arises whether such a test-object can be used for testing the capacity of the lens to resolve details of other types, or whether we must seek other test-objects approximating more closely to those which arise in practice. In the latter case the above theory must be used only as an approximate indication of general trends. There is already some evidence in the literature that the value of resolving power obtained for a photographic material varies with the form of test-object employed. Ross (1924) found that a test-object consisting of a number of parallel lines separated by a distance equal to their width gave the same values of resolving power as a " double-star " test-object consisting of pairs of disks, but higher values than those given by a fan-shaped test-object. Sandvik (1927) investigated the variation of resolving power with the ratio of width

of line to width of space, and found that the resolving power for a given photographic material increased uniformly with the logarithm of the line/space ratio. In 1928 he showed experimentally that the resolving power increases with the contrast of the test-object. The theory to explain these two results put forward by Frieser (1935, 1938) was based on the sine-curve test-object, and it was through his work that the simplification introduced by using a sine-curve test-object was first appreciated.

Some early experiments were made with a test-object (shown in figure 3) which is fairly typical of the parallel-line type. This was photographed by means of an 8-inch F/2·9 lens stopped down to F/8, using different films, at different exposure levels and different focus settings. A study of the negatives showed that the length of the lines of the test-object influenced the observer's estimate of the smallest group which could be resolved. For example, there was usually a gradual transition from groups which could be recognized clearly to groups which could not be recognized, so long as the resolution limit occurred at some group within a set of five of constant line length. When, however, the transition from resolved to unresolved groups occurred at a group near X in figure 3, where there is a sudden change in line length, it was found that the larger group on one side of X was clearly defined, whilst the smaller group on the other side of X showed no trace of line structure. The number of lines also influenced the result, for if only a small portion of the group at the resolution limit were exposed to view, the observer did not recognize it as resolved, but when more of the area was exposed to view the observer had no hesitation in deciding that the pattern was recognizable. The results obtained with the test-chart described by Sayce (1940), in which the lines are all of the same length but with the separation between them diminishing from end to end of the chart, are not likely, therefore, to bear any simple relation to those obtained with other charts.

Experiments were also carried out using letter charts of the type used by ophthalmic opticians, for the reason that the letters would presumably be unknown to the observer, and thus provide a better test of recognition than test objects the shape of which was already known to the observer. Such letter charts were, however, not very satisfactory, for they did not provide a very uniform scale of detail, and in any case the observer soon memorized the sequence of letters. There seemed little doubt from the somewhat scattered evidence that different resolving powers would be found with different test-objects, but this would be of no importance if the resolving powers yielded by different test-objects always bore the same ratio to each other. To check this point, comparisons were made between the parallel-line test-object shown in figure 3 and a test-object which was designed to eliminate the defects of the parallel-line type. This new object contained the minimum number of lines necessary for a determination of resolving power (namely, two) and in which the ratio line-length/line-width was constant at all sizes. Moreover, the length of line was no greater than was necessary to permit the recognition of the longer dimension, and therefore the direction, of the test-object. A test-object described by Cobb and Moss (1928), consisting of two bars of length three times their width, separated by a distance equal to their width, had hitherto been used for visual acuity measurements; its dimensions fitted the above desiderata sufficiently closely to suggest the name " Cobb test-object " for this type. The experiment was carried out with a composite chart consisting

of groups of Cobb test-objects and groups of parallel line test-objects. Photographs were made at different settings of focus of an 8-inch F/2·9 " Pentac " lens, and the limit of resolution was assessed by a number of observers. The resolving-power limit as determined by the Cobb test-object was plotted against that determined by the parallel-line test-object, as in figure 4. Each point represents the mean estimate of a number of observers. (The departure of some of the points from the smooth curve is probably due to the fact that in the parallel line test-object the line frequencies of consecutive groups were in the ratio of 1·2 to 1, whereas in the Cobb test-object the ratio was 1·1 to 1.) It will be noted that although at the position of focus the two test-objects give substantially the same value of resolving power, there is an appreciable departure from equality on each side of the focus ; at positions inside the focus, the parallel-line test-object leads to higher
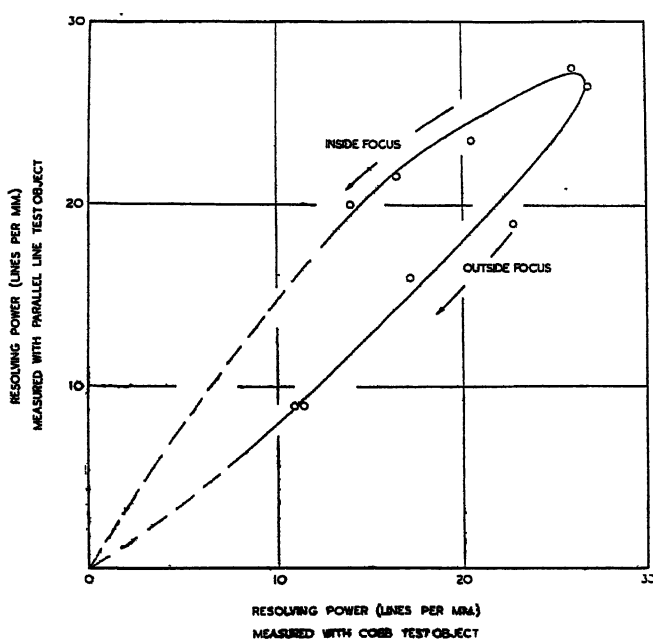
Figure 4. Comparison of parallel-line and Cobb test-objects.

values of resolving power than the Cobb test-object, whilst beyond the focus the reverse is true. The explanation for these results must be that the distributions of light in the image given by the lens are not the same on both sides of the focus position. The effect of this change must then be different in the image of the multi-line test-object from that in the image of the two-line test-object.

Obviously the resolving powers for different test-objects do not, in general, bear the same ratio to each other, and if resolving-power tests are to be of value, the test-object must be chosen to simulate, as near as may be, the type of objects to be photographed. It was therefore decided to adopt the Cobb type of test-object, which is probably as nearly similar to the detail in air photographs as can be obtained in a simple geometrical pattern. Although no comparisons with other test-objects for accuracy have been made, it is possible that this type of test-object does not yield results of quite so high precision as others. By using a number of Cobb test-objects disposed at random, and plotting the number of a given size
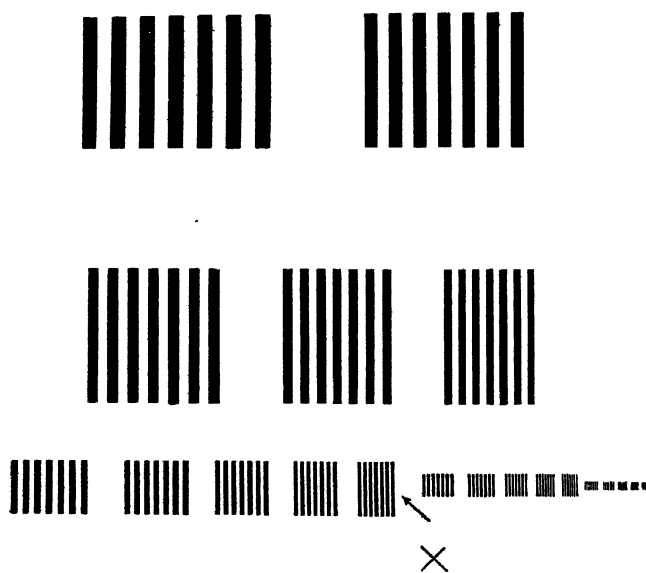
Figure 3.   Parallel-line test-object.

Figure 5.   Form of test-object adopted for resolution tests.

recognized against the size an estimate may be made of the size at which half of each group are recognized, and the precision may be thereby somewhat improved (Romer and Selwyn, 1943). This method could not be applied to lens examination mainly because the complete test-object must be kept reasonably small. In the form used for the present work the test-chart consists of groups in a spiral arrangement, as shown in figure 5. Half the test-objects are vertical and half horizontal, so that measurements can be made along radial lines and along tangential lines. The definition of the lens used in preparing the test-chart and of the collimator used for the tests is likely to be worse for points off axis, so the smaller groups, the sharpness of which will be relatively more affected by poor definition in these two lenses, are placed near the centre.

Aerial photographs are different from ordinary photographs in that the atmosphere itself appears luminous and throws a veiling light over the surface of the earth, as a result of the scattering of the light by small suspended particles, Thus while an average subject photographed from the ground has a brightness range from the brightest part of the subject to the darkest of 160 : 1 (Jones, 1941), aerial photographs show that the ratio for subjects seen from moderate and high altitudes is only 10 : 1 or so. Detail can range in ratio of maximum to minimum brightness from 1 : 1 to 160 : 1 for photographs taken on the ground, and to 10 : 1 for aerial photographs. A rough mean for the detail in the terrestrial photograph may therefore be taken as $\sqrt{160} : 1$, and for the aerial photograph as $\sqrt{10} : 1$, say 3 : 1. Examination of details in aerial photographs larger than that just resolved, so as to avoid the reduction in contrast due to lack of sharpness, suggested that a much more nearly typical value of ratio of brightness was 1·6 : 1. In terms of the usual photographic units, therefore, the appropriate density difference in the test-object was estimated at 0·2 (i.e. log 1·6). Figure 5 has been reproduced to approximately this density. It is possible that for testing lenses for use in ordinary photography on the ground a somewhat higher contrast than this is desirable.

Suppose we now return to the example shown in figure 1, and consider what happens when the contrast is decreased. To decrease the contrast of a point source we must place around the point source an extended source of uniform brightness (as occurs for instance in astronomy where the stars are surrounded by the general illumination in the night sky). The image given by the lens then becomes as shown in curve B of figure 1 (*a*) and the photographic image as in curve B of figure 1 (*b*). It will be seen that when the contrast is low the photographic image of the point source extends to the full diameter of the optical image. Thus we may conclude that changes in the exposure are likely to affect the image less if the contrast is low and that the outer and weaker parts of the optical image are important when the contrast is low, but may be unimportant if the contrast is high. The latter point is, in fact, usually true whether the image is observed visually or photographically. Suppose that the images of a point source given by two lenses have intensity distributions as shown in figure 6 at A and B. A represents an image of moderate diameter, into which all the light from the lens comes to a focus. B represents a central bright image surrounded by a flare of 1/100 of the intensity, covering an area 100 times that of the bright core; the amount of light in the core is thus equal to the amount of light in the flare. The log P_t curves corresponding to these images are shown at A′ B′ of figure 6. The visual and photographic

threshold lines are annotated "Visual High Contrast" and "Phot. High Contrast". If the contrast of the point object is now reduced by surrounding it by a uniformly illuminated field, the log P_l curves A' and B' should be shifted downwards by an appropriate distance but, for convenience, the threshold lines have been shifted upwards by the same amount. They are annotated "Visual, Low Contrast" and "Phot. Low Contrast" respectively. In order not to complicate the argument, the effect of the spread of light in the emulsion layer is neglected. It will be seen that the resolving power when the contrast is high is determined very largely by the bright central core of the image which suffers from flare, but that as the contrast is decreased the flare becomes much more important. We shall be referring to this point later in connection with the effect of aperture on the resolving power.

Choice of photographic materials

It has already been mentioned that the prospects of simplifying the problem by using the theory of resolution were remote. The work therefore had to be
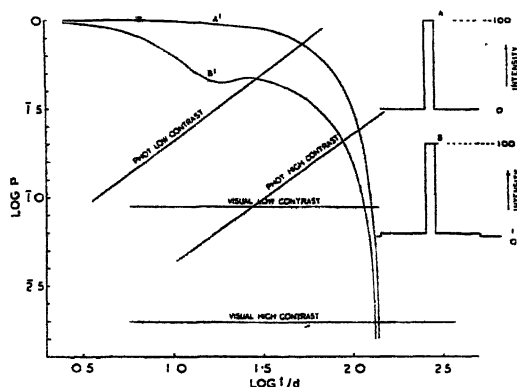
Figure 6. Effect of flare round image on log P_l curves, and effect of this on resolving power for high and low contrast test-objects.

proceeded with on an empirical basis, and for this reason the choice of photographic materials was governed by two practical considerations: firstly, it was required to know what standard of performance was obtained with existing equipment, and secondly, it was required to know how far this standard might be expected to rise, if the properties of the film could be improved, or if by some means the exposure time could be lengthened to permit the use of slower fine-grain materials. The first consideration naturally led to the choice of the existing Service type films, the Kodak versions of which were "Super-XX Panchromatic Aero Film" and "Panatomic X Aero Film". The former is the faster of the two and has a lower resolving power. Partly in order that the effects of chromatic aberration of the lenses should remain the same in all the photographic tests, and also because panchromatic sensitization was not likely to be superseded by any other, such as infra-red, it was desirable that the film chosen to satisfy the second requirement should also be panchromatic. It should also have high resolving power and moderate speed. No existing high-resolution film has high speed, but the emulsion manufacturer is continually striving to increase resolving power whilst retaining a useful speed. Allowing for possible advances in this direction,

and also in devices (for reducing vibration and compensating for the forward motion of the aircraft) which would permit of longer exposure times, it was decided to choose Kodak "Microfile" Panchromatic film. This has a resolving power about three times that of Super-XX, whilst its speed is about 1/80 that of the latter. If the speed of this film could be increased, and the exposure time increased, there was some possibility of it being brought into use.

Development conditions

Since the resolving power of the film is dependent on the contrast, not only of the test-object, but also of the emulsion, it was important that development conditions should be standardized. The Service type films were developed to give about the same contrast as obtained in the R.A.F. Photographic Reconnaissance Units, namely $\gamma^* = 1\cdot4$ for Super-XX and $\gamma = 1\cdot6$ for Panatomic-X, both films being developed in Kodak developer D.19b. Microfile film, in its normal use for microcopying documents, etc., is developed to a high contrast, but for pictorial photography a high contrast in the negative would lead to a density range too great to be printed on existing positive materials. Development was therefore standardized to make $\gamma = 1\cdot6$. Kodak developer DK.20 was used initially, but this was later replaced by Kodak developer D.76d, which was more conveniently available, and conditions were adjusted to produce the same negative characteristics as formerly.

Description of apparatus

The examination of lenses of focal lengths from 5 inches to 20 inches was carried out with the apparatus shown diagrammatically in figure 7. The test-object (B) was mounted at the focus of a collimator objective (A), and illuminated by light from a tungsten filament lamp. An opal diffusing screen (C) to ensure uniform illumination, and a minus-blue filter (as used in R.A.F. practice) were
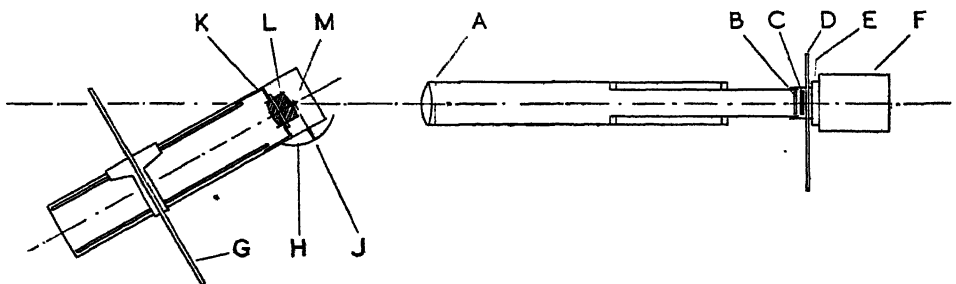
Figure 7. Arrangement of apparatus.

mounted between the source and the test-object. The test-object was exposed by means of a roller blind shutter (E) attached to the lamp-house (F) and a slide (D) carrying neutral filters, graded so that the brightness of the test-object could be altered in steps by a factor of $\sqrt{2}$, was fitted between the lamp-house and the shutter. The collimator objective was a Wray achromatic doublet of focal length 60 inches and aperture F/15, permitting lenses of aperture 4 inches to be tested.

* In this connection γ is the slope of the straight portion of the characteristic curve of the material.

The camera incorporated a lens holder (K), capable of carrying lenses of external diameters up to 6 inches, and a film holder (G) sliding on rails, guided by two V-grooves and a knife-edge. The film (in strips 35 mm. wide) was held against a register glass of thickness 0·15 inch, in conformity with the standard register glass used in the smaller R.A.F. cameras. The length of the film holder was such that a total angle of view of 56° could be covered with a 20-inch lens. The camera could be rotated about a vertical axis M the position of which was adjusted to coincide with the front surface of the lens to be tested. This was to ensure that the lens did not move out of the beam of light from the collimator.

A larger collimator and camera was constructed later to accommodate lenses of focal lengths up to 50 inches. The collimator lens was a Cooke " Photo Visual " objective of focal length 134 inches and aperture 7 inches.

Both collimator objectives were well corrected, and their influence on the results obtained with different aircraft lenses was negligible. The lateral aberrations of a collimator objective of focal length F are reduced by a factor f/F in the image plane of a test lens of focal length f, and the longitudinal aberrations by a factor f^2/F^2. A number of 20-inch lenses, examined with the aid of both collimators in turn, showed the same values of resolving power, within the limits of experimental error.

§3. EXPERIMENTAL PROCEDURE

Exposure conditions

The illumination in the focal plane of a lens is not uniform, but decreases as the angular separation from the axis is increased. In a thin lens the effective aperture is reduced as the angle of incidence is increased, and the principal ray strikes the focal plane at a greater angle the greater the angle of incidence. Thus the intensity of illumination I at any angle is given by $I = I_0 \cos^4 \theta$, where I_0 is the intensity of illumination at the centre of the field. If the lens is of finite edge thickness, or, more particularly, consists of separated components, an additional decrease in illumination occurs, due to "vignetting". Absorption and reflection losses due to the register glass, as used in R.A.F. cameras, also increase with angle of incidence. Minor variations with angle arise from reflection and absorption losses in the different components.

The exposure given to the film therefore decreases with increasing angular separation from the axis. Since the resolving power of the film varies with the exposure it receives, the resolving power of the lens and film used together will vary with angle due to film properties as well as to lens properties. In addition, the subject may cover a considerable range of brightnesses: in aerial scenes this range is comparatively small, as stated above. In order to make a complete survey of the problem with regard to these practical considerations, it would be necessary to collect data on lens-film resolution for a number of different exposure levels. It was thought, however, that such a procedure would unnecessarily lengthen the programme of experimental work and complicate analysis of the data, and all variations in resolving power for a given lens-film combination were restricted to those due to the lens alone, by adopting one exposure only. The exposure chosen was that which would result in the maximum film resolution which occurs, for the most negative materials, when the density of the negative is about 1·0.

Measurements of the relative illumination at different positions in the field of each lens were made in order to discover how serious the variation was, and whether the technique of giving constant exposure to all parts of the film was justified.

Measurement of relative illumination in the focal plane

The relative illumination in the focal plane of each lens was measured by photographic photometry in the following way. Using an illuminated disk in the place of the test-object, exposures were made, of constant duration, at 0° and at 5° intervals on each side of the axis, up to the full angular extent of the test-lens, on a film of moderate contrast. This was repeated for different apertures of the lens, and the exposure level for each aperture was adjusted so that the densities obtained on development lay on the linear portion of the sensitometric curve of the material.* Sensitometric film-strips were developed together with the test-strips exposed by the lens, and the values of log exposure corresponding to the densities produced by the lens were read from the sensitometric curve thus obtained. Since we require to know only the relative values of the intensity at different points in the field, these log exposure values may be considered to be values of log intensity.

In order to allow for possible variations in the shutter speed, and the intensity of the light source, two exposures were made (by moving the film laterally inside the film-holder) at each position, and, therefore, at all except the axial position there were four exposures for each angle. It was found that there was little variation in the densities so obtained, and the values of log intensity were usually reproducible to within about ± 0.02. Greater precision than this could have been obtained by controlling lamp-voltage and shutter speed, but was considered unnecessary.

Measurement of visual resolving power

Although the research was to be primarily for the purpose of assessing the photographic performance of lenses, preliminary experiments had shown that some useful information might be gained if the resolving power of the aerial image were measured. The design of the apparatus did not permit visual measurements to be made easily for extra-axial images, and these measurements were confined to the axial image. Test-objects were made on Kodak "Maximum Resolution" Plates, to the same design as for the photographic test-object. The plates referred to were the only type available for recording faithfully the number of lines per mm. required for the visual measurements. Two test-objects were made, one having clear lines on an opaque background, and the other having clear lines on a background of density 0·2. The high contrast test-object was used in order to measure the resolving power of a lens in the way in which this measurement is usually carried out, and the low contrast test-object to provide measurements of resolving power approaching more closely those obtained photographically.

The test-objects were placed at the focus of the collimator objective, and a microscope was used to observe the image produced by the lens under test. The focus of the microscope was adjusted until the maximum number of groups of the test-object was resolved, and the number of the smallest resolvable group was recorded. The relative aperture of the microscope objective was always greater than that of the lens under test.

---

* It is not essential to the method that all the densities on the test negatives lie on the linear portion of the characteristic curve of the material, but this ensures the same degree of accuracy for all values.

Photographic resolution tests

After preliminary trials to ascertain what region of focus was to be explored, and to find the exposures necessary to give a density of about 1·0 at each position in the field, the film-holder was set and exposures were made on the axis and at 5° intervals up to the full angular extent of the lens, on both sides of the axis. This procedure was repeated at different settings of the film-holder, until seven planes, passing through the region of focus, had been explored. This series was then repeated at all apertures of interest, and for the other films.

Since most of the lenses were examined at about 5 apertures, this procedure resulted in about 100 negatives per lens, with 20 or more images on each negative. The amount of work involved in measuring the resolving power in the negatives, not to speak of the complexity of the problem of analysing the data, imposed limitations on the thoroughness with which the inspection could be carried out. It was therefore necessary to limit the measurements to one negative at each aperture, for each of the three films.

For this reason, from each set of 7 negatives, that one was chosen in which the best axial definition was obtained. The only other plane of "focus" of any significance would correspond to the negative in which the best average definition was obtained, having regard to the entire area of film which the lens was intended to cover. The selection of this plane would involve the inspection of all the images, · however, and would not solve the problem. This method was used later in connection with the focusing of lenses in cameras. It is worth pointing out that in a lens afflicted with spherical aberration both the plane in which the best axial definition occurs and the optimum focal-plane shift away from the lens as the aperture is reduced. Thus in setting cameras to give the best performance at different apertures without re-focusing, it is necessary to compromise on a focal setting which will be the optimum with respect both to the range of apertures and to the angle of view to be covered. Having chosen the negative, the number of the smallest group of the test-object recognizable at each angle was recorded and converted into resolving power in lines per mm., applying a correction discussed below. The resolving powers for radial and tangential lines were recorded separately.

In outline the method is similar to that adopted by Gardner and Case (1937) at the National Bureau of Standards, but there are several points of divergence between the two methods. The National Bureau of Standards apparatus has a number of fixed collimators accurately positioned so that measurements of distortion may be made. A high contrast test-object is used in which the number of lines and the ratio of length to separation is not the same for all line frequencies. The photographic material used is the Eastman "Type VB Spectrographic" plate, which has the same order of resolving power as "Microfile" film. The problem in hand at the National Bureau of Standards is to calibrate lenses in terms of the optimum focal length having regard both to resolution and distortion, and is, therefore, different from the problem with which this paper is concerned. ·

Measurement of resolving power in negatives

In determining the limit of resolution from the negative, the conditions of viewing were first standardized. The illumination of the field of the microscope was kept fairly constant from negative to negative, and the degree of magnification

was kept constant for a given photographic material within the obvious limits that the magnification should not be so small that the size of test pattern is near the limit of the acuity of the eye, nor so large that the obscuring power of the graininess obliterates the image. It was found that a convenient magnification was 20 for the Service type films, and this had to be increased to 75 when observing good images on "Microfile" film. Since all the negatives were of approximately the same density, the conditions of illumination were kept constant, once a satisfactory level of illumination had been found. It is an advantage of the low contrast test-object that glare is absent when viewing the negative. The "spurious" resolution effect referred to by Gardner and Case (1937) has not been observed by the authors with the low contrast Cobb test-object. This effect, which is observable with multi-line high contrast test-objects, is that lines of a particular size appear to be resolved, although closer inspection reveals more lines than appear in the original test-object. With the Cobb test-object one should observe a strong line in the middle and possibly weaker lines on the outside if this effect occurred, but in fact, when the contrast of the test-object is low, the effect cannot occur.

The measurement of resolving power is to some extent subjective and is liable to errors which may appear large by physical standards. It is usually necessary for the observer to set himself some criterion of resolution, with the result that some observers judge as resolved a pattern which other observers would consider inadequately defined. These differences occur largely amongst inexperienced observers, and it has been found that when a number of workers have gained experience the agreement between them is good. The resolving-power determinations were carried out in the present work by one observer, who adopted the criterion that a group of a given size was resolved if the Cobb element appeared to consist of two separate patches, each *quasi* rectangular in form, and if the direction in which the "lines" were oriented agreed with that on the original test-object. It is sometimes found that the pattern appears to consist of two lines pointing in a direction at right angles to that in the original test-object. This is due to fortuitous "clumps" of grain.

L. C. Martin and C. A. Padgham* have investigated the precision of photographic resolution measurements under conditions similar to those described in this paper, and their results agree with estimates made by the present authors which suggest a standard deviation of about 10% for repeated measurements on the same negative. It was found, in addition, that the percentage precision of measurement was the same for both high and low contrast test-objects.

In converting the number of the smallest groups which can be resolved into resolving power in lines/mm., a correction must be applied for the increase in magnification for extra-axial images. In practice the resolving power was calculated from the dimensions of the test-object rather than by measurements on the negative. If $1/R_0$ is the separation between lines in the object, F the focal length of the collimator, f the focal length of the test lens, and θ the angular separation from the axis at which the image falls, then the line frequency R in the negative is given by

$$R_R = (R_0 F \cos \theta)/f$$

and
$$R_T = (R_0 F \cos^2 \theta)/f,$$

where the suffixes R and T indicate radial and tangential lines respectively.

* Private communication.

As a result of this distortion of the test-object the length of radial lines becomes greater than 3 times the width of line whilst the length of tangential lines becomes less than 3 times the width of line by a factor $\cos\theta$. This change of shape is, however, small enough to be of little importance.

§4. EXPERIMENTAL RESULTS

The lenses examined in this survey were mainly of British manufacture, although a few American and German objectives were included. The focal lengths ranged from 5 to 49 inches and their maximum apertures from F/2·9 to F/7. The types of construction included wide and narrow angle, triplet and telephoto types but in spite of this variety there was sufficient similarity in their performance to enable results to be calculated as typical of the majority for presentation below.

Variation of resolving power with aperture

In figure 8 resolving power is plotted against $F.No.$, both on logarithmic scales, for six of the lenses tested. The results are quite typical. The measurements refer to the axis only, but include visual measurements with high and low contrast test-objects, as well as photographic measurements on the three films with low contrast test-objects. The broken line indicates the value of resolving power calculated from the formula $1/\lambda(F.No.)$, λ being taken as 0.55μ. This is a customary value to adopt for a perfectly corrected lens (Conrady, 1929). Strictly, it refers to visual observation of two close point-sources seen against a non-luminous background. It will be seen that except at the highest apertures the observed resolving powers for high contrast Cobb test-objects approximate to those given by the formula, being usually some 10% lower. The greater departures at the higher apertures must be due to uncorrected aberrations. With low contrast test-objects the visual resolving powers are inevitably less than for high contrast test-objects, but the deviations from the formula at large apertures are also so considerable that the curves of resolving power against $F.No.$ display a definite maximum. The photographic resolving powers are much lower than the visual, and diminish in the order Microfile, Panatomic-X, Super-XX. They also display maxima in the curves of resolving power against $F.No.$ and the maxima tend, though perhaps not in strongly marked fashion, to larger values of $F.No.$ in the order Microfile, Panatomic-X, Super-XX.

It is not difficult to account for these results in approximate fashion. To do this we shall use the theory already mentioned, which, although strictly applicable only to sine-curve test-objects, may be expected to give a qualitative account of the results to be expected from other types of test-object. It is first necessary to make some estimate of the curves of $\log P_l$ against $\log(1/d)$ for different apertures, and for this purpose it is convenient to note that if the lens were perfectly corrected (and if the light from the test-object were monochromatic—though this does not make an important difference) the curves for $\log P_l$ for different apertures would be all of the same shape but separated from each other in the $\log(1/d)$ direction by an amount equal to the logarithm of the ratio of successive F-numbers. A set of such curves is shown by the light lines of figure 9.

When the path differences introduced by imperfect correction are less than half a wave-length, the diameter of the central nucleus in the image of a point source is no greater than when correction is perfect. There is, however, some
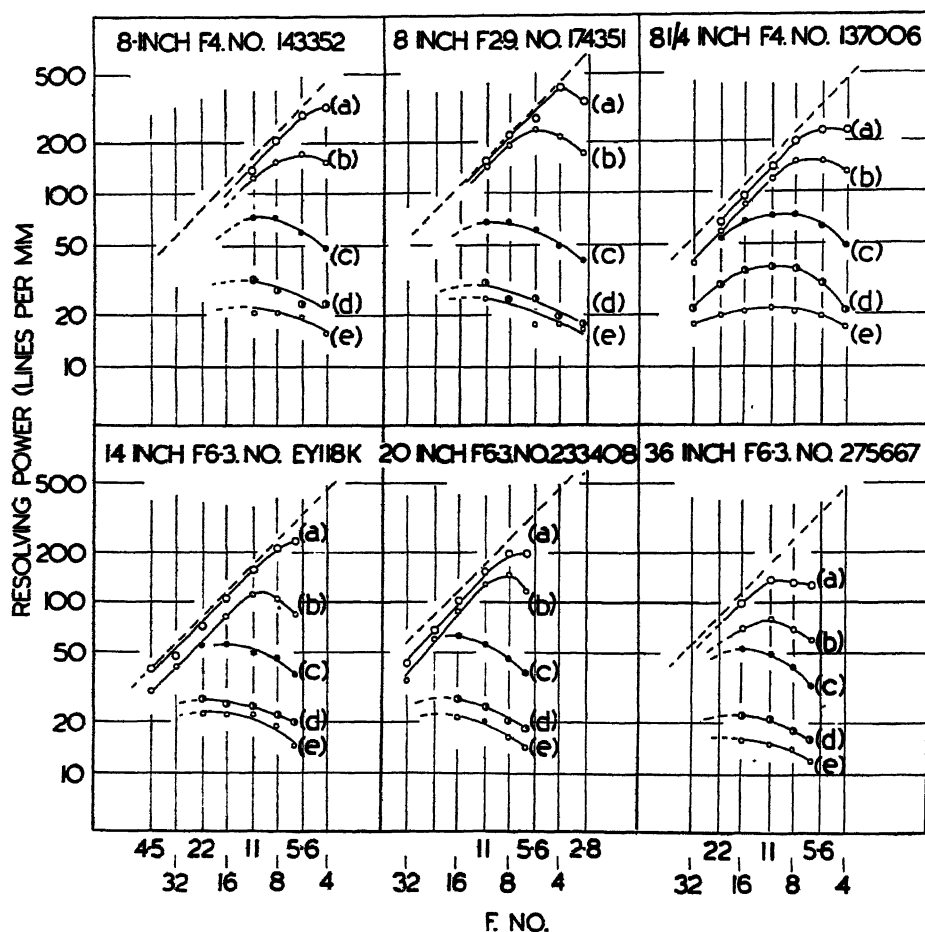
Figure 8. Typical curves showing variation of visual and photographic resolving power with aperture (for axial region only).

(a) Visual, (b) Visual, (c) Microfile Film, (d) Panatomic X film, (e) Super XX film.
Curve (a) refers to high contrast test-object, and curves (b) to (e) to low contrast test-object.
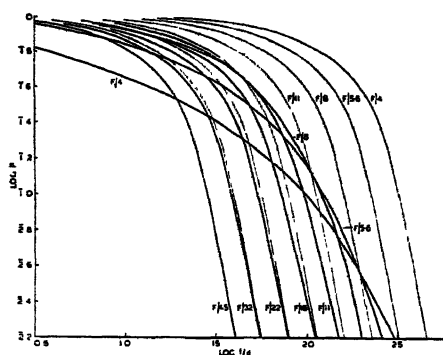
Figure 9. Log P_l curves for perfectly and imperfectly corrected lens at different apertures. Thin lines are curves for perfectly corrected lens and thick lines for imperfectly corrected.
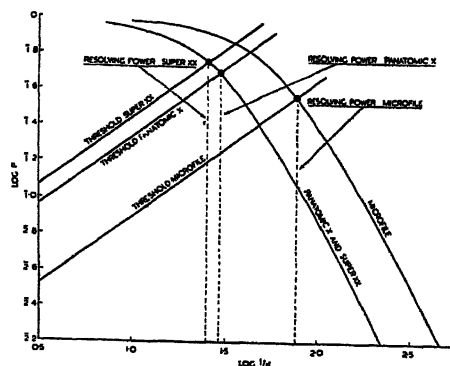
Figure 10. Estimation of log P_e curves from resolving power of films.

34-2

loss of light, and the light which is lost from the nucleus appears in the rings surrounding the nucleus, usually as a general luminous halo. (Martin, 1926; Conrady, 1929). An illustration of the effect of a halo of this sort is provided by figure 6, which may thus be used as a guide for estimating how the curves in light lines of figure 9 must be modified when a small amount of aberration is present. When the amount of aberration is large the central nucleus, if it still persists, becomes larger than in an aberration-free image at the same aperture. The nucleus may not be evident in some cases, but there also the effective size of the image is greater than that of the corresponding aberration-free image. Thus when the amount of aberration is considerable the $\log P_l$ curve necessarily lies substantially to the left of the curve for a perfectly corrected lens of the same aperture, even at low values of $\log P_l$. When the amount of aberration is small, on the other hand, as indicated above, the $\log P_l$ curve remains of approximately the same shape as that for a perfectly corrected lens, but is shifted downwards slightly.

Now the general result of decreasing the aperture of a lens is to reduce the path differences at the focus. Thus at high apertures the $\log P_l$ curves will, if the aberrations are considerable, lie substantially to the left of the curves shown by light lines in figure 9, while for small apertures they will not differ much from these latter curves, but will tend to lie a little below them. To add more precision it has been assumed that at F/45 the lens is practically free from aberration, while at F/4 a guess at the shape of the curve has been based on the values of the visual resolving power obtained on lens No. 137006 for test-objects of high and low contrast. The curves for the apertures intermediate between F/4 and F/45 were drawn by eye to form a consistent family. The whole family of curves thus obtained is shown by heavy lines. It should be noted that a family of real $\log P_l$ curves might not display such an obviously systematic variation as those shown in figure 9.

We now require to estimate the influence of the diffusion of light in the emulsion layer. Fairly satisfactory results·have been obtained, in other connections, by assuming that the intensity of illumination in the emulsion decreases exponentially with the distance from a long, narrow, brightly illuminated strip of the emulsion. If this is true, then the curve of $\log P_e$ against $\log (1/d)$ has a definite shape which is the same for all photographic materials but which lies towards higher values of $\log (1/d)$ the less the diffusion of light in the emulsion layer. The actual position of the P_e curve can be located on the $\log (1/d)$ scale, once the threshold line (determined by the granularity and γ-value of the material) has been fixed, and the resolving power measured. This is illustrated in figure 10. The resolving powers of the three films were measured using a contact test-object, with a density difference of 0·2 between the light and dark parts. It should not be concluded that the diffusion of light is exactly the same in Panatomic-X as in Super-XX—the calculation is not a very exact one.

Adding together the $\log P$ curves of figures 10 and 11 gives the curves of $\log P_l + \log P_e$, and these are shown in their upper range in figure 11 for Super-XX and Panatomic-X, for which they are conveniently identical. Figure 11 also shows the curves of $\log P_l$ over the lower range: for low values of $\log P$ only visual resolution is possible, and it is not necessary to include the $\log P_l + \log P_e$ curves. The resolving power at any aperture may be read from these curves. When plotted, the variation of resolving power with aperture is as shown in figure 12, which also includes values for Microfile film. The qualitative resemblance to the

curves shown in figure 9 will be obvious, and in the main phenomena shown by photographic lenses at the centre of the field may be regarded as accounted for. It
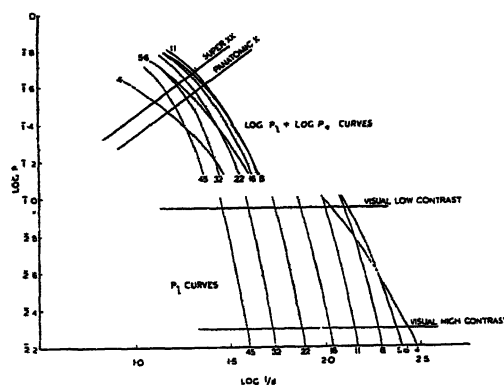
Figure 11. Log P_l and log P_l+log P_e curves and threshold curves for imperfect lens shown in figure 9.
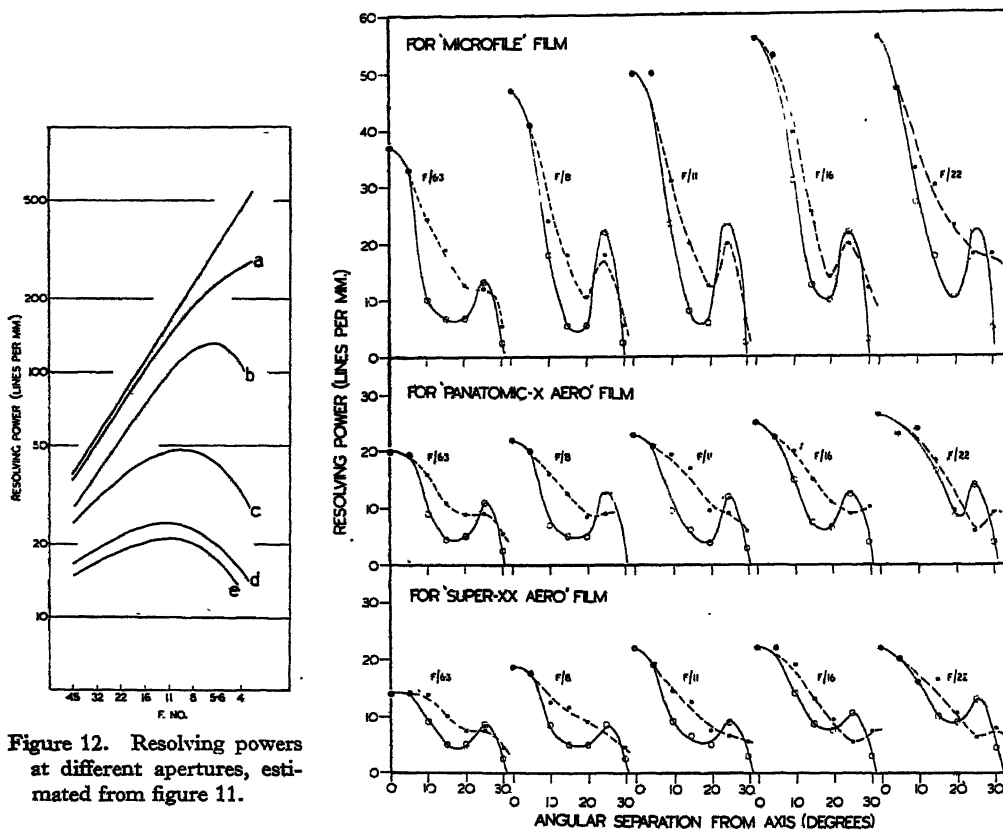
Figure 12. Resolving powers at different apertures, estimated from figure 11.

Figure 13. Typical [resolving power, angular separation from ax curves. Full curves: radial lines. Dotted curves: tangent lines.

should be particularly noted that the results of the tests are inconsistent with any assumption except that at full aperture important amounts of aberration exist at the centre of the field.

Although there is no published evidence that the variation of photographic resolving power with aperture had been investigated, either theoretically or experimentally, before the work described in this paper was commenced, there has been some discussion in photographic literature of the fact that the resolving power of a lens when used with photographic materials is much less than the figure usually accepted as the theoretical limit. As the discrepancy is particularly great at high apertures, it has been suggested on several occasions (for example, Zeigler, 1936) that the reduction in resolving power is due to the wide angle of the cone of light incident on the film and the finite thickness of the emulsion layer. In a paper seen late in the progress of this work Roeder (1941) also subscribes to this view, although he does, in addition, suggest that at high apertures the lower contrast of the image has a greater effect upon the photographic than on the visual resolving power. The proposed explanation is, however, not correct, as has been shown by Gregory (1946). A similar explanation for the poor resolving power observed off-axis has also been shown by Gregory to be incorrect. Pinoir (1944) also discusses the above points, and shows that there is serious quantitative disagreement between the proposed explanation and the observed resolving power. It is interesting that in the paper quoted, Pinoir discusses the phenomenon of resolving power by arguments very similar to those we have used. His work was, of course, quite independent of ours.

Variation of resolving power with angle

Curves of resolving power against angular separation from the axis (hereafter referred to as (R, θ) curves) were obtained for these lenses in the three films. In figure 13 is a set of such curves for the 14-inch lens No. EY 118K. The (R, θ) curves for each aperture refer to the plane in which the best axial resolving power occurs at that aperture. The general character of the curves in figure 15 is typical of all the lenses so far examined. Individual lenses show minor divergencies of shape, but the similarity in the (R, θ) curves over a wide range of types of lens is surprising. At the particular focus setting adopted, the resolving power is highest at the centre of the field. At points off-axis there is, except at special points, a difference between the resolving powers for radial and tangential lines, and the resolving-power curve for radial lines usually has a secondary maximum at some point in the filed.

The same general shape of curve is obtained for all three films, the average level of resolving power being lower when the Service films are used than when Microfile film is used. This is to be expected because of the greater granularity and the greater diffusion of light in the Service films. But it has already been shown that aberration of an important amount exists in the centre of the field, where the resolving power is greatest. Consequently it must be concluded that, away from the centre, aberration is present to an even more important degree. Stopping down the lens is for that reason more effective in improving photographic resolution at points off-axis than on-axis, as may be seen from figure 13. Some curious changes, such as in the curves for Microfile film, where the minimum at F/8 is lower than at F/6·3 but increases as the aperture is decreased below F/8, are almost certainly due to the re-focusing at each aperture.

One curious feature of the (R, θ) curves for radial lines deserves some mention. It was found that the secondary maximum and the minimum in the (R, θ) curves

occurred at more or less constant positions in the field. Table 1 gives the tangents of the angles corresponding to these positions, and also the tangents of the semi-angle of view, defined as the angle at which the resolving power falls to zero. It will be seen that, with few exceptions, the maximum occurs at about 0·8 of the linear field and the minimum at about 0·5 of the linear field.

In most photographic objectives, the field curvature is such that the sagittal and tangential surfaces depart from the Gaussian image plane, and then re-cross it at some point towards the edge of the field. It is to be expected, therefore, that if the shape of the (R, θ) curve is determined by field curvature, the resolving power should first decrease as the angular separation from the axis is increased from zero, and then increase as the image surface swings back to re-cross the "focal plane".

Table 1. Positions of secondary maximum, minimum and zero in (R, θ) curves for radial lines

Lens		tan θ			Ratio (1) : (3)	Ratio (2) : (3)
		At max. (1)	At min. (2)	At zero (3)		
30-inch	156449	0·18	0·09	0·29	0·62	0·31
20-inch	73051	0·36	0·25	0·49	0·73	0·51
20-inch	154107	0·27	0·18	0·36	0·75	0·50
20-inch	233408	0·47	0·32	0·58	0·81	0·55
14-inch	227628	0·38	0·28	0·49	0·78	0·57
14-inch	EY118K	0·47	0·32	0·58	0·81	0·55
10½-inch	263403	0·47	0·28	0·58	0·81	0·48
8¼-inch	137006	0·58	0·36	0·73	0·79	0·49
8-inch	230863	0·47	0·28	0·58	0·81	0·48
8-inch	174351	0·47	0·32	0·58	0·79	0·55
5-inch	143352	0·58	0·36	0·75	0·77	0·48

Means...... 0·77   0·50

Suppose that the surface upon which radial lines are imaged (the sagittal surface) is generated by the rotation about the x-axis of the curve

$$x = \alpha y^2 - \beta y^4. \qquad \ldots\ldots(1)$$

The secondary maximum in the (R, θ) curve will then occur where the above curve crosses the focal plane, that is at

$$y_2 = (\alpha/\beta)^{\frac{1}{2}}, \qquad \ldots\ldots(2)$$

and the secondary minimum where x is a maximum, that is at

$$y_1 = (\alpha/2\beta)^{\frac{1}{2}}. \qquad \ldots\ldots(3)$$

The "edge" of the field cannot be determined easily. However, there is a fairly close limit to the "edge" since the curve of x falls away from the focal plane extremely rapidly after crossing it. It may be assumed that worse definition is tolerated at the "edge" of the field than at the secondary minimum in the (R, θ) curve. That is to say, a greater value of x may be tolerated at the "edge" of the field than at y_1, and the limit of the field may be determined by making x at the "edge" $-\eta$ times as great as at y_1, where η is greater than unity. Thus

$$\eta\alpha^2/4\beta = \alpha v_o^2 - \beta v_o^4.$$

whence
$$y_3 = \left[\frac{\alpha\{1 + (1+\eta)^{\frac{1}{2}}\}}{2\beta} \right]^{\frac{1}{2}}.$$

Thus the ratio y_2/y_3, which corresponds with the ratio (1):(3) of table 1, is given by

$$y_2/y_3 = \left[\frac{2}{1 + (1+\eta)^{\frac{1}{2}}} \right]^{\frac{1}{2}}.$$

and the ratio y_1/y_3, which corresponds with the ratio (2):(3) of table 1, is given by

$$y_1/y_3 = \left[\frac{1}{1 + (1+\eta)^{\frac{1}{2}}} \right]^{\frac{1}{2}}.$$

The first ratio is equal to 0·75 and the second to 0·53 when η is equal to 6, which is probably not an unreasonable value.

The ratio y_2/y_1 is independent of any assumption as to the value of η, and is equal to $\sqrt{2}$. The experimental value, as determined by the ratio of the η values for the secondary maximum and the minimum, is 1·56, i.e. within 10% of the calculated value. Considering that no special efforts were taken to ascertain the precise positions of these features, the agreement is close enough to give some support for the assumption that for these lenses the sagittal field curvature is of the form described by equation (1). It may be observed that the *relative* positions of the minimum, secondary maximum, and zero in the (R, θ) curves are independent of the coefficients α and β. Equation (1) represents the third and fifth order field-curvature for an infinitely narrow pencil of rays. The positions of the secondary maximum and minimum do not change when the aperture is varied, however, and it may be inferred that the above discussion is equally valid for lenses of high aperture.

In the published field-curves for a number of lenses (for example, Hay and von Rohr, 1932) the tangential field-curves display the characteristics of equation (1) as often as do the sagittal field-curves, and yet the (R, θ) curves for tangential lines rarely exhibit the secondary maximum. The reason for this may be that the aberrations of aperture, which do not apparently affect the validity of the above discussion for radial lines, are more serious in their effect on the resolving power for tangential lines. Such defects as coma and transverse chromatic aberrations, as well as the effects of vignetting, might be expected to exert a greater influence on the resolution of tangential lines than of radial lines.

It has already been mentioned that the drop in resolving power away from the axis cannot be explained away as a result of the obliquity of the light. Washer (1945) has recently shown that even if a lens is completely free from aberrations the resolving power will be less for points off-axis than on-axis, owing to the smaller angular aperture of the cone of light reaching the image. The effect is, however, too small to be of any importance in real objectives of the type tested, and in any case cannot be put forward to account for the present results since reducing the aperture improves the definition.

Variation of illumination across field

Typical curves of log relative illumination plotted against angular separation from the axis are shown in figure 14. In most cases the measurements were made up to angles much greater than the lenses were intended to cover. The decrease

in illumination from centre to corners of the film size for which the lenses were designed was in most cases fairly small. The resolving power of most photographic materials increases rapidly as the exposure is increased from zero, and reaches a maximum value when the mean density on the film is about 1·0. At greater exposures the resolving power falls slowly. If the exposure at the centre of the field is such as to give a density of 1·0, at other points in the field the film will have a lower effective resolving power if no compensation is introduced. In practice the variation of film-resolving power may be reduced to a minimum by arranging that the optimum exposure is reached at some intermediate zone of the film.
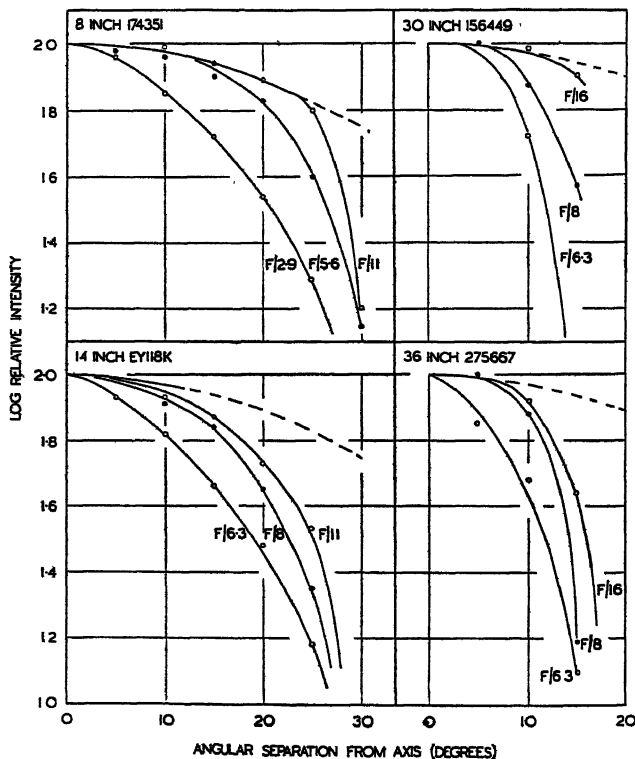
Figure 14. Decrease of intensity towards corners of field.
Broken curve represents $I = I_0 \cos^4 \theta$.

From a curve of film-resolving power against log exposure, the effective film-resolving power at the corners of the field was calculated for the standard R.A.F. film sizes, and it was found that in only a few lenses was the vignetting serious enough to introduce any appreciable change in the effective film-resolving power. Moreover, since at the corners of the field the definition is determined by the lens rather than the film, the resolving power of the lens-film combination is very insensitive to changes in film characteristics.

Although vignetting tends to reduce the effective film-resolving power, it may have a beneficial use in removing some of the aberrant light from outer zones of the lens at the higher obliquities. Examples of this effect were encountered in studying the results of the present work.

§5. ANALYSIS OF RESULTS

An exact but simple comparison between the performances of different lenses is almost impossible. Not only is the problem complicated by differences in focal length, aperture, angle of view, and the differences in resolving power for radial and tangential lines, but the variation of resolving power with angle from the axis is different for different lenses. If, however, we are prepared to forego detailed comparison, and to take a more general view, the problem is not so difficult. The interpreter of an aerial reconnaissance photograph requires to be able to recognize details of objects on the ground. In practice this means, at present standards of performance, that the finer the detail resolved the better the photograph. The value of a photograph is also dependent upon the number of details visible in it, that is upon its area. If, therefore, a system of calculation can be devised which will give a figure for the total number of resolved details in an aerial photograph, that figure may be used as an indication of the performance of the camera in question. Some considerable discussion has taken place on this point and the best system of calculation is not yet agreed upon. We shall in this account keep to the original method.

If the resolving powers for radial and tangential lines are R_R and R_T respectively, we can imagine that this represents the resolution of a rectangular unit of area $1/R_R R_T$. Consider the plane of the film to be divided up into a series of concentric zones by means of circles of radii r_0, $r_0\sqrt{2}$, $r_0\sqrt{3}$, etc., so that each zone has area πr_0^2. The number of resolvable elements in a zone is evidently $\pi r_0^2 \bar{R}_R \bar{R}_T$, if $\bar{R}_R$ and $\bar{R}_T$ are the mean values of the resolving powers over the zone. The values of r_0 were chosen somewhat arbitrarily to be 10 mm. for a lens of 5 inches focal length and proportionately for other focal lengths, so that corresponding zones for all lenses fall at the same angular separation from the axis. It follows that the number of resolvable units in any zone is proportional to $(Rf)^2$, where R is the geometric mean of $\bar{R}_R$ and $\bar{R}_T$. The figure is thus proportional to the square of the mean angular resolving power.

The calculation of the averages was facilitated by re-plotting the (R, θ) curves as (R, n) curves, where n is the number of zones enclosed by an angle θ to the axis and is equal to $161 \cdot 3 \tan^2 \theta$.

Having thus arrived at a method of estimating the performance of a lens, on a given film, we are now in a position to examine whether there is any general rule connecting the performance of a lens with its aperture, focal length, angle of view and type of film with which it is used.

There were indications of such a general rule before the full data were obtained. For instance Hay and von Rohr (1932) give figures for spherical aberration, coma and field curvatures for a large number of lens specifications, taken from patents. These are all for a standard focal length of 100 mm. The curves given for spherical aberration extend to a value of aperture which is presumably the full aperture at which the lens is intended to work, and the curves for curvature of field extend to an angle which is presumably the angle covered by the lens. When these values are plotted one against the other and a mean curve is drawn through them it is found that there is an approximately constant ratio between the F. No. and the square of the angle of view. This is a natural consequence of the difficulty of reducing aberrations to tolerable values at high apertures and wide angles. On

the assumption, therefore, that all these lenses were intended to give approximately the same level of performance, it may be concluded that the relation between the average number of resolvable units per zone, $\overline{N}$, the $F.\,No.$, the number of zones, n_∞, which the lens covers before the resolving power falls to zero, and the focal length, f, will be approximately of the form

$$\frac{F.\,No.}{n_\infty} = \phi(\overline{N}, f),$$

since

$$\theta_\infty^2 \doteq \tan^2\theta_\infty = n_\infty/161\cdot3.$$

In an attempt to determine the nature of ϕ, a three-dimensional model was constructed, in which the average number of resolvable units per zone (obtained on Super-XX film) was plotted against $(F.\,No./n_\infty)$ and f. This model confirmed that $(F.\,No./n_\infty)$ was an appropriate argument in the functional relation, and furthermore, showed that the $F.\,No.$ need not be restricted to the maximum aperture. The same relation held, approximately, for all apertures. To test whether this was also true if n was varied, the average number of resolvable units per zone up to the nth zone was plotted against $F.\,No./n$ for different apertures. With one or two exceptions, a single curve was obtained for each lens ; in the exceptional cases, different values of $\overline{N}$ were obtained for the same value of $F.\,No./n$, according to whether the $F.\,No.$ were high or low. In general, however, the agreement was good enough to warrant further consideration, and on re-plotting the curves on double-logarithmic paper, the points were found to approximate to a straight line for each lens. At this stage the data for the other films were included and a number of graphs, such as that shown in figure 15, were obtained.

Assuming that the points for different apertures lie in the same straight line, the average number of resolvable units per zone $(\overline{N})$ may be expressed in the form

$$\overline{N} = A\left[\frac{F.\,No.}{n}\right]^b, \qquad \dots\dots(4)$$

where A is a function of the focal length and the film properties and b is the slope of the straight lines and varies, though slightly, from lens to lens and film to film. The value of b is approximately 0·6. If, on the other hand, it is necessary to account for the fact that in some cases the points for different apertures lie on different straight lines, equation (4) must be written

$$\overline{N} = A\frac{(F.\,No.)^b}{n^c}. \qquad \dots\dots(5)$$

For the sake of simplicity, we shall adopt the form in equation (4), for we can thus set $F.\,No./n = 1$ and find a single value of A for each lens-film combination. This value of A was found to be directly proportional to focal-length, in the following manner :—

> for " Super-XX " film $A = 3kf$,
> for " Panatomic-X " film $A = 4kf$,
> for " Microfile " film $A = 15kf$.

The factor multiplying the focal length contains a constant k the value of which depends upon the units in which the resolving power is measured and on the size of the zones into which the film is divided.

The following considerations may help to explain why $\bar{N}$ is proportional to f. Suppose we have two series of lenses of the same $F. No.$ but of different focal lengths, such that in one series there are no aberrations whatever, and the other series is of lenses all of the same design, but made to different scales, so as to have different focal lengths, and with considerable amounts of aberration. For the lenses without aberrations the resolving power in lines/mm. will be the same for all focal lengths, because the image size is independent of the focal length. The angular resolving power will therefore be proportional to the focal length and
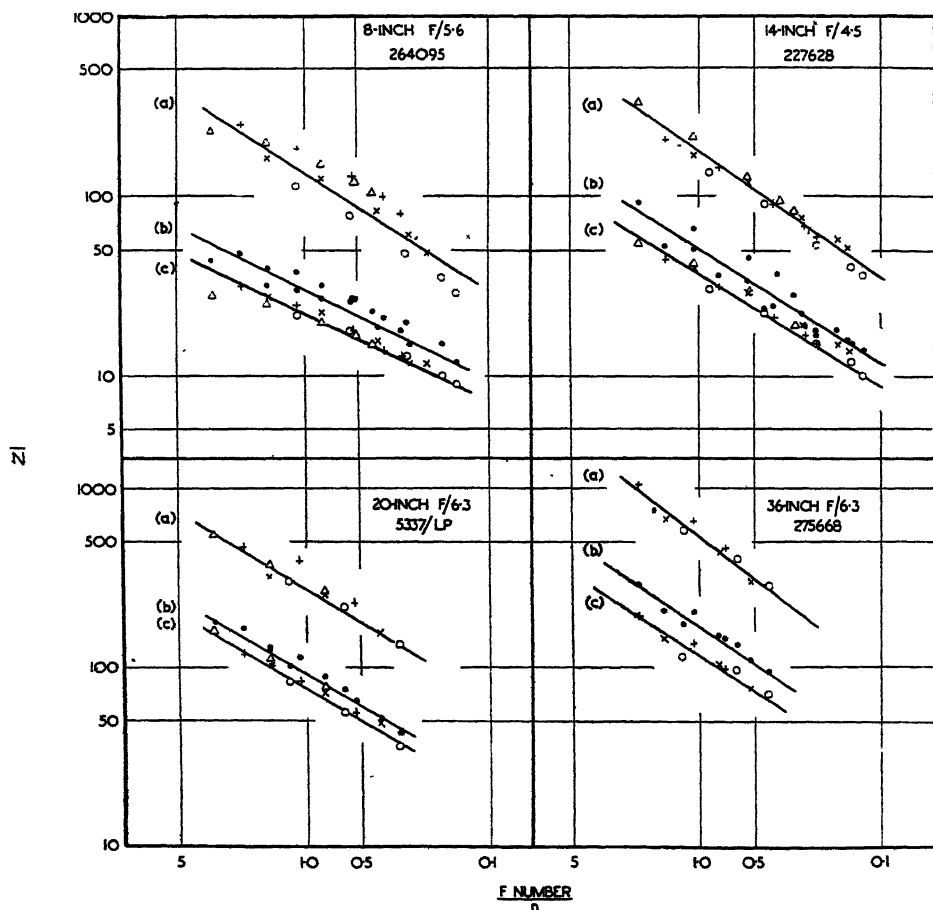
Figure 15.   Variation of $\bar{N}$ with $F. No./n$.
(a) Microfilm film, (b) Panatomic-X film, (c) Super-XX film.

$\bar{N}$ will be proportional to the square of the focal length. When the lenses have considerable amounts of aberration, the size of the image is proportional to the focal length, so that the angular resolving-power, and $\bar{N}$ also, will be independent of focal length. In fact, even in this case, $\bar{N}$ will increase a little with focal length because the diffusion of light in the emulsion layer will be of less importance when the optical image is large, and also because the threshold of resolution is lower for large images than for small. Now the path-length errors are proportional to the focal length, and, therefore, however well or ill the lens may be designed,

there is some short focal length at which the lens is nearly perfect and some long focal length at which the correction is bad. In other words, for short focal lengths $\bar{N}$ will vary as f^2 and for long focal lengths will be constant, or nearly so. For moderate focal lengths, therefore, it is not unreasonable to find $\bar{N}$ varying approximately as f. What is perhaps more surprising is that the proportionality between $\bar{N}$ and f is not seriously disturbed by changes in aperture. Figure 16, which is illustrative only, will perhaps make the argument clear.

The general level of the resolving power varies according to the film used for taking the photographs. Two quantities have arisen during the theoretical discussions to which these differences of level might be attributed. They are the granularity and the diffusion of light in the emulsion layer. The dimensions of granularity are [Density: Length] and the dimension appropriate to the diffusion of light is [Length], since it may be conveniently measured by the distance in which the light flux in the emulsion layer coming from a long illuminated narrow strip drops by the ratio e. Either quantity, therefore, will satisfy the dimensional
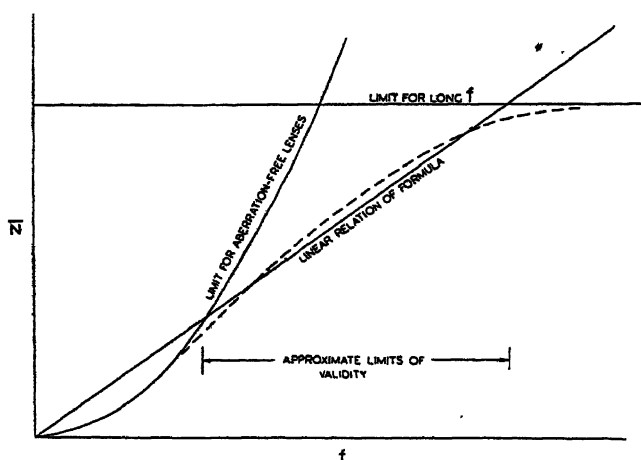
Figure 16. To show that for short focal lengths $\bar{N} \propto f^2$, for long focal lengths $\bar{N}$ is constant, and that intermediately $\bar{N} \propto f$, approximately.

requirements of the formula. But in fact the diffusion of light is not likely to be of any importance. It will be seen from figure 2 that log P_e does not drop much below 0 for values of $1/d$ less than 20. Values as high as this are observed only on the axis. As soon as any appreciable angle is covered, the resolving power drops considerably below 20 lines/mm. and the diffusion of light in the emulsion layer is an unimportant item in the degradation of the image.

The granularity values, measured by the method described by Romer and Selwyn (1943), at a density of $1\cdot0$ are $1\cdot6$, $1\cdot2$, and $0\cdot3$ density-microns for Super-XX Aero, Panatomic-X Aero and Microfile films respectively. The reciprocals of these figures are in the ratio $3:4:16$, from which it may be concluded that the complete formula is

$$\bar{N} = \frac{Kf}{G}\left[\frac{F.\,No.}{n}\right]^{b}, \qquad \ldots\ldots(6)$$

in which b has the approximate values of $0\cdot55$, $0\cdot6$ and $0\cdot75$ for Super-XX Aero, Panatomic-X Aero and Microfile films respectively. Substituting these values

in (6) and comparing the calculated value of $\bar{N}$ with the experimental values, K is found to be approximately $5 \cdot 5 \times 10^4$.   In table 2 the departure of the performance of individual lenses from that of the "average" lens as given by (6) is shown, and in table 3 the frequency with which these differences occur is shown.

Table 2.   Mean percentage differences of experimental values of $\bar{N}$ from those given by equation (6)

Lens \ Film	Super-XX ($b=0\cdot55$)	Panatomic-X ($b=0\cdot6$)	Microfile ($b=0\cdot75$)
5-inch F/4 143352	−21	−10	− 3
8-inch F/4 230863	−14	−13	+ 9
8-inch F/2·9 174351	−12	− 8	− 2
8-inch F/5·6 264095	−15	− 6	+ 7
8¼-inch F/4 137006	− 7	+21	+30
10½-inch F/6 263403	+ 4	+ 3	− 2
14-inch F/4·5 227628	−24	−24	−24
14-inch F/6·3 EY118K	− 5	− 8	−12
14-inch F/5·6 72754	+25	+27	+19
20-inch F/5·6 73051	+16	+23	+15
20-inch F/5·6 277490	−19	..	..
20-inch F/6·3 154107	− 3	− 8	−20
20-inch F/6·3 174856	+10	..	+50
20-inch F/6·3 5337/LP	+10	− 1	−16
50 cm. F/5 2451477	−10	..	− 1
24-inch F/6 7887	− 3	..	− 5
24-inch F/6 EA504	− 2	..	−32
30-inch F/6·3 156449	+34	+19	+69
36-inch F/6·3 275667	− 9	− 1	−14
49-inch F/7 402490	−23	−28	−39 .

Table 3.   Distribution of differences in $\bar{N}$.

Percentage difference	No. of cases		
	(Super-XX)	(Panatomic-X)	(Microfile)
10% or less	10	8	7
11%–20%	5	2	6
20% or more	5	5	6

The standard deviation for repeated measurements of the resolving power of the same image on the same negative is of the order of 10%, but in view of the large number of readings made for a given lens, at different angles to the axis, and since all the readings were made by the same observer, the averaged values of resolving power are probably reliable to within much less than 10%.   (It should be noted that since $\bar{N}$ is proportional to the *square* of the angular resolving power, the differences in terms of angular resolving power are approximately half of the values listed in table 2.)   The differences shown in table 2 may therefore be regarded as real differences, and due to differences in performance between different lens types.

A single value of b may be adopted without serious change in the reliability of the empirical relation. Thus, eliminating the parameters introduced originally,

$$\bar{N} = 5 \cdot 5 \times 10^4 \times \frac{f}{G} \left[\frac{F.No.}{n} \right]^{0 \cdot 6} \qquad \text{resolvable units per zone}$$

$$= 2 \cdot 6 \times 10^3 \times \frac{f}{G} \left[\frac{F.No.}{\tan^2 \theta} \right]^{0 \cdot 6} \qquad \text{since } n = 161 \cdot 3 \tan^2 \theta$$

or
$$\frac{2 \cdot 6}{4\pi} \times \frac{10^3}{f^2} \frac{f}{G} \left[\frac{F.No.}{\tan^2 \theta} \right]^{0 \cdot 6} \qquad \begin{array}{c} \text{resolvable units per sq. mm.} \\ \text{on the film} \end{array}$$

$$= \frac{207}{fG} \left[\frac{F.No.}{\tan^2 \theta} \right]^{0 \cdot 6} .$$

The mean resolving power on the film is thus

$$\bar{R} = \left[\frac{207}{fG} \right]^{\frac{1}{2}} \left[\frac{F.No.}{\tan^2 \theta} \right]^{0 \cdot 3} \text{ lines per mm.}$$

and the mean angular resolving power, in radians^{-1}, is

$$25 \cdot 4 \left[\frac{207f}{G} \right]^{\frac{1}{2}} \left[\frac{F.No.}{\tan^2 \theta} \right]^{0 \cdot 3}, \qquad \qquad \cdots \cdots (7)$$

since f is measured throughout in inches.

§6. DISCUSSION OF EMPIRICAL RELATIONSHIP

The formula which has been arrived at to represent a general view of the performance of aircraft camera lenses has no fundamental or theoretical basis to ensure that it is generally valid. This has been made plain during the latter part of·this investigation, as a result of measurements on new designs. These have shown some improvement in performance, which, while it does not so far make the above formulae seriously in error, undoubtedly shows that with continued improvement a new formula, approximating more closely to the theoretical limit of perfection, will be necessary. For the time being, however, the above formulae are a satisfactory statement of the level of performance to be expected from a lens of given dimensional properties, provided that these dimensional properties are not outside the range of those of the lenses tested.

To illustrate this last point it is convenient to use again the argument of §5. It will be obvious that the rule $\bar{N} \propto f$ cannot be obeyed for very short focal lengths for which $\bar{N} \propto f^2$, so that the empirical formula predicts higher resolving powers than are theoretically possible. And it will also be obvious that $\bar{N} \propto f$ cannot be true beyond focal lengths at which the lens may be regarded as badly corrected.

The correction of aberrations is better in the centre of the field than towards its outer parts, with the result that the diffusion of light in the emulsion layer is relatively more serious in the centre of the field than elsewhere. Thus the importance of the diffusion of light in the film is dependent upon the angle of view. In the formula the influence of diffusion does not appear explicitly, but may be involved implicitly in the term G, since there is an approximate connection between G and the diffusion. But the influence of granularity also varies over the field of the lens, as a result of its relatively greater obscuring effect for fine detail than for

large. To represent the influence of granularity by a single term unconnected with the angle of view is, therefore, theoretically incorrect. It is more than probable that the different values of b found for the three materials arise because of this effect. Unfortunately the data are quite inadequate for investigating this question more fully, and any attempt at more precise formulation of the experimental data than is represented by equation (5) would be somewhat artificial.

There are a few minor limitations to the generality of the results. The average resolving power was computed for circular films, since the analysis would have been rather complex if rectangular films had been considered. In addition, in the collimator method of examining lenses, only a small fraction of the total field of the lens is illuminated, and the effect of light scattered by the lens is less than under practical conditions. Another departure from strictly practical conditions was that the brightness of the test-object was increased for extra-axial exposures to compensate for the decrease in illumination due to vignetting, etc.

In spite of the foregoing discussion, it is somewhat surprising, in view of the known differences in construction of the lenses under examination, that their performances may be defined, as well as they are, solely by the dimensional properties of focal length, aperture and angle of view. The empirical formula implies the existence of a standard of performance, and this can only mean that the designer endeavours to reduce the aberrations to a level which experience has taught him will lead to a satisfactory result. Another designer may aim for a different type of correction, but the extent of the correction upon which he decides will probably be such as to lead to a photographic performance approximately equal to that accepted by the first designer. This is not to say that the empirical expression is inflexible; the lenses upon which it was based were designed during and after the 1914–18 war and up to the early years of the second war, and all that is claimed is that the formula represents the standard prevailing up to about 1943, when this work was completed. Indeed, since that time advances have been made which have raised the performance of individual lenses to such an extent that a repetition of the present research with a wide enough range of the latest products of the British optical industry would result, if not in a change in the form of the relationship, then certainly in a change in the constants. For the time being, however, the formula is useful, not only for its original purpose, but also as a standard of reference by which the merit of a given lens for air photography may be estimated by comparing its performance with that calculated from the formula, and for estimating the effect of moderate changes in the focal length, aperture and angle of view of a lens of given basic design.

§ 7. ACKNOWLEDGMENTS

REFERENCES

Cobb, P. W. and Moss, F. K., 1928. *J. Franklin Inst.*, **205**, 831.
Conrady, A. E., 1929. *Applied Optics and Optical Design* (Oxford).
Frieser, H., 1935. *IX Congrès Internat. de Phot.*, Paris, 207.
Frieser, H., 1938. *Z. wiss. Phot.*, **37**, 19.
Gardner, I. C., and Case, F. A., 1937. *J. Res. Nat. Bur. Stand., Wash.*, **18**, 449.
Gregory, J. M., 1946. *Proc. Phys. Soc.* (in course of preparation).
Hay, A., and von Rohr, M., 1932. . *Handb. wiss. und angew. Phot.*, Band 1, " Das Phot. Objektiv" (Vienna : Julius Springer).
Jones, L. A., 1921. *J. Opt. Soc. Amer.*, **5**, 232.
Jones, L. A. and Condit, H. R., 1941. *J. Opt. Soc. Amer.*, **31**, 651.
Martin, L. C., 1926. *Trans. Opt. Soc., London*, **27**, 249.
Pinoir, R., 1944. *Sci. Industr. Photogr.* (2), **15**, 257.
Roeder, H., 1941. *Phot. Industr.*, **39**, 385, 401, 418, 432, 439.
Romer, W. and Selwyn, E. W. H., 1943. *Phot. J.*, **75**, 571.
Ross, F. E., 1924. *The Physics of the Developed Photographic Image.* Eastman Kodak Monograph, No. 5.
Sandvik, O., 1927. *J. Opt. Soc. Amer.*, **14**, 169 ; 1928. *Ibid.*, **16**, 244.
Sayce, L. A., 1940. *Phot. J.*, **80**, 454.
Washer, F. E., 1945. *J. Res. Nat. Bur. Stand., Wash.*, **34**, 175.
Zeigler, H. W., 1936. *Amer. Phot.*, **30**, 553.

THE DETERMINATION OF DISSOCIATION ENERGIES BY THE BIRGE-SPONER EXTRAPOLATION

By A. G. GAYDON,

Warren Research Fellow of the Royal Society, Imperial College, London

ABSTRACT. A critical examination is made of the Birge-Sponer method of determining dissociation energies of diatomic molecules. Dissociation energies of certain classes of molecules as determined by linear extrapolation are compared with values determined independently. For the ground states of ordinary molecules the linear extrapolation tends to give too high results, often by around 20%. For molecules in which ionic binding is important the extrapolation often comes too low, sometimes much too low. For the molecules in which one atom has a $^1S_0$ ground state the linear extrapolation is much too high. The method appears to work satisfactorily for ionized molecules. Excited states of molecules cannot, in general, be relied on to give good extrapolations. These empirical observations are discussed in terms of the potential-energy curves and the binding forces in the molecules at large inter-nuclear distances.

---

§ 1. INTRODUCTION

THE method of determining the dissociation energies of diatomic molecules by extrapolation of the vibrational energy levels to their convergence limit was first put forward by Birge and Sponer (1926). Since then some other methods have been used as well to ascertain the dissociation energies of certain molecules, but for the large majority of diatomic molecules we are still dependent on the Birge-Sponer extrapolation.

When data are available for a large number of vibrational levels, the graphical

extrapolation, made by plotting the vibrational energy intervals $\Delta G(v+\frac{1}{2})$ against the vibrational quantum number v, or sometimes against the vibrational energy term $G(v)$, may be expected to lead to a fairly reliable result. In the majority of cases, however, only a few levels are known and the interval between these decreases approximately linearly, and it is then customary to make a *linear extrapolation*. It has been generally realized that there is a tendency for this linear extrapolation to come too high. Birge (1929) discussed the form of the curve of $\Delta G(v+\frac{1}{2})$ against v for a few molecules and showed that its shape and curvature changed rather unexpectedly in some cases ; the sudden inflections, which he believed he had established for the excited states of O_2 and some of the halogens, have not, however, received much support from later, more precise, spectroscopic observations. Rydberg (1931) suggested that an alternative form of extrapolation, in which $[\Delta G(v+\frac{1}{2})]^2$ is plotted against v, might be better in some cases ; and since this method usually gives a lower value of the dissociation energy than that obtained by linear extrapolation, it has sometimes been used. Nevertheless, the linear extrapolation is still in most general use and forms the basis for most of the values of dissociation energies now in the literature.

For a fair number of molecules accurate values of the dissociation energy have been obtained by other methods, such as by observation of band structure up to the convergence limit at which continuous absorption begins (O_2 and the halogens), of atomic fluorescence (halides of alkali-metals, Tl, etc.), of predissociation (HgH, N_2, CO, SO, etc.), or by thermochemical methods (alkali-metal halides, etc.) or by combinations of these methods (HCl, HBr, HI, NO). These methods have been reviewed in the author's recent book (Gaydon, 1946) and the values for the dissociation energies have been listed. It is the purpose of this paper to make a systematic comparison between these accurate values of dissociation energies with those obtained by linear extrapolation, with a view to making better use of the linearly extrapolated values for the many molecules for which this is the only method available.

In their original paper Birge and Sponer pointed out that their method of extrapolating was only justified if the vibrational energy levels converged to a limit at a finite value of the vibrational quantum number v. Kratzer (1924) showed that if the force between the nuclei at very large internuclear distance, r, is expressed as a power series in $(1/r)$, then, if the dominant term at large r is $1/r^2$ or $1/r^3$, there will be an infinite number of vibrational energy levels, so that the curve of $\Delta G(v+\frac{1}{2})$ against v would approach the v axis asymptotically. If, however, the dominant term in the expression for the force is $1/r^4$ or higher power of $(1/r)$, then the maximum value of v will be finite. For non-polar molecules the force at large internuclear distance varies as a high power of $(1/r)$ and the extrapolation is therefore permissible. For polar molecules dissociating to ions, however, the force must vary as $1/r^2$ at large distances and the extrapolation should not be made by the method of Birge and Sponer. It is clear that the rate of convergence of the vibrational levels near the limit will depend on the form of the law of force at large distances and we may, therefore, expect the curve of $\Delta G(v+\frac{1}{2})$ against v to show a different trend for classes of molecules with different forms of binding. It therefore appears best to treat separately certain general classes of molecules.

§2. EXTRAPOLATION FOR THE GROUND STATES OF NEUTRAL MOLECULES WITH NON-IONIC BONDING

First, we shall compare the accurately determined dissociation energy D_0 and the linearly extrapolated value D_{lin} for the ground electronic states of those neutral molecules for which ionic binding forces are not unduly important. The values of the dissociation energies are taken from the author's book, where full references may be found. For N_2, NO, CO and CN the values have been discussed by Gaydon and Penney (1945) and differ from those in the older literature (e.g. Herzberg, 1939). For SO and S_2, values higher than those given by Herzberg have been adopted.

It is possible to make the linear extrapolation either graphically from the first part of the curve of $\Delta G(v+\frac{1}{2})$ against v or analytically. The latter method is more usual and has been adopted here.

If the energies of the vibrational levels are expressed as a power series

$$G(v) = \omega_e(v+\tfrac{1}{2}) - x_e\omega_e(v+\tfrac{1}{2})^2 + \dots, \qquad \dots\dots(1)$$

where v takes integral values 0, 1, 2, ..., it can be shown that, neglecting higher terms, the energy interval from the lowest level, $v=0$, to the limit, which is the required dissociation energy , is

$$D_{lin} = \frac{\omega_e^2}{4x_e\omega_e} - \tfrac{1}{2}\omega_e. \qquad \dots\dots(2)$$

It is often given as $D_{lin} = \omega_0^2/4x_e\omega_e$ where $\omega_0 = \omega_e - x_e\omega_e$; for all practical purposes this is satisfactory, but it may be pointed out that it is not strictly correct as it differs from (2) by $\frac{1}{4}x_e\omega_e$.

The values of D_{lin} have been calculated with the aid of equation (2) from collected values of ω_e and $x_e\omega_e$ given by Herzberg (1939) ; for ClF the value is from a more recent paper by Wahrhaftig (1942). Table 1 shows the values of D_0, D_{lin}, and the ratio D_0/D_{lin}.

Table 1. Values of D_0, D_{lin} (in ev.) and D_0/D_{lin} for the ground states of molecules with non-ionic bonding

Molecule	D_0	D_{lin}	D_0/D_{lin}	Molecule	D_0	D_{lin}	D_0/D_{lin}
H_2	4·478	4·55	0·99	O_2	5·084	6·3$_0$	0·81
KH	1·86	2·02	0·92	SO	5·148	6·3	0·82
AgH	2·3	2·7	0·85	S_2	4·4	5·7	0·77
OH	4·4	5·3	0·83	Se_2	2·8	4·5	0·62
				TeO	2·725	5·7$_6$	0·48
HCl	4·431	5·15	0·86				
HBr	3·60	4·78	0·75	N_2	9·764	11·8	0·83
HI	2·75	4·0	0·69	P_2	5·033	6·6$_6$	0·76
				As_2	3·94	5·1	0·77
Li_2	1·12	1·46	0·77	Bi_2	1·72	2·27	0·76
Na_2	0·77	1·07	0·71				
K_2	0·51	0·74	0·69	ClF	2·616	1·93	1·35
				Cl_2	2·476	2·44	1·02
CO	11·11	11·2	0·99	ICl	2·153	3·08	0·70
NO	6·49	7·6	0·85	Br_2	1·971	3·02	0·65
CN	7·6	9·9	0·77	IBr	1·818	2·85	0·64
				I_2	1·542	2·37	0·65

The table includes practically all the non-ionic molecules for which adequate data for D_0 and $x_e\omega_e$ are available. The exceptions are certain molecules containing elements of Group II of the Periodic Table, which have $^1S_0$ ground states and which, for reasons given presently, are best considered separately. It will be noted that OH, HCl, HBr and HI are included in table 1 ; ionic forces probably contribute to a certain extent to the bonding ·in these molecules, but since the electron affinity of the non-metal is in each case small (around 3 to 4 ev.) compared with the ionization potential of hydrogen (13·5 ev.), these molecules will dissociate to neutral atoms and the force between the atoms at large distances is unlikely to be appreciably affected by ionic forces.

It will be seen from table 1 that there is a general tendency for the true dissociation energy to lie below the linearly extrapolated value. The average value of D_0/D_{lin} is 0·79, and for thirteen of the 28 molecules the values lie between 0·75 and 0·85, with six above 0·85 and nine below 0·75. Thus we see that for this class of molecules the most probable value of the dissociation energy is about 20% below that given by the linear extrapolation. Clearly the scatter in the values of D_0/D_{lin} is rather great, but apart from the general tendency for D_{lin} to be too high it is possible to distinguish at least one systematic effect. D_0/D_{lin} tends to be highest for molecules composed of light elements and lowest when heavy elements are involved. This effect is well shown both by the group containing the halogen molecules and by that consisting of molecules built from Group VI elements. It is not easy to express this effect in quantitative form, but it would seem that one might do worse than to adopt a dissociation energy 10% below the linearly extrapolated value for hydrides and molecules composed of light elements, while for molecules containing a very heavy element a value only 70% of that given by linear extrapolation seems reasonable.

It may be pointed out that if the experimental data are not of high accuracy there may be a tendency for the next term, $y_e\omega_e(v+\tfrac{1}{2})^3$, in the series expansion for $G(v)$ (equation (1)), to be masked by the experimental error. Since this term is usually negative, the result will be to obtain a slightly higher value of the apparent $x_e\omega_e$ than the true one, thus bringing D_{lin} down slightly.

§ 3. MOLECULES CONTAINING AN ATOM WITH $^1S_0$ GROUND STATE

The form of the graphical Birge-Sponer extrapolation or the accuracy of the linear extrapolation must clearly be related to the potential-energy curve for the molecular state, especially its shape at fairly large internuclear distances. Now an atom with a closed shell or closed sub-shell of electrons, i.e. an atom in a $^1S_0$ state, has no resultant electric or magnetic moment and is, therefore, likely to exert a smaller force on another atom at a distance than is an atom with a resultant moment. This is borne out by the fact that the rare gases do not, with the exception of some excited states for He_2, form molecules at all. For the elements of Group II of the Periodic Table, however, although their ground states are $^1S_0$, many stable diatomic molecules are known. In these cases the binding forces are believed to be mainly due to excited states of the atom, although as a result of interaction between the various potential-energy curves and the rule that curves of the same species may not cross, the ground states of these diatomic molecules do probably correlate with ground atomic states. Clearly, the form

of the extrapolation is likely to be affected by the fact that the binding is largely due to excited states, and that at large internuclear distances the force between the atoms is likely to be abnormally small. It thus seems desirable to treat molecules of this type separately.

Although a large number of diatomic molecules formed from elements of Group II are known, for relatively few of them is it possible to compare D_0 and D_{lin}; apart from the scarcity of accurate values of the dissociation energy, for many of these molecules, especially the halides, the band systems are not strongly degraded, so that the values of $x_e\omega_e$ derived from the analyses of band heads are far from accurate. The data are most complete for the hydrides; for many of these predissociation has been observed, which sets at any rate an upper limit for D_0. The values are given in table 2.

Table 2. D_0, D_{lin} (in ev.) and D_0/D_{lin} for hydrides of Group II elements

Molecule	D_0	D_{lin}	D_0/D_{lin}
ZnH	0·84	1·35	0·62
CdH	0·678	1·29	0·53
HgH	0·372	0·64	0·58
BeH	~2·3	3·6	~0·64
CaH	≤ 1·70	2·6	≤ 0·65
SrH	≤ 1·66	2·6	≤ 0·64
BaH	≤ 1·82	2·6	≤ 0·69

Thus for these hydrides the true value of the dissociation energy is much less than the linearly extrapolated value, often being only about 60% of that value. The unusually rapid rate of convergence of the vibrational levels for these molecules is shown by the high values of $x_e\omega_e$ and the strong negative curvature of the graphs of $\Delta G(v+\frac{1}{2})$ against v in many cases. For molecules other than hydrides the evidence is scanty. For CaCl predissociation shows that $D_0 \leqslant 2·76$, while a poor linear extrapolation gives 5·1 ev., so that $D_0/D_{lin} \leqslant 0·54$. For HgCl recent measurements (Wieland, 1941) give a linear extrapolation to 1·6 ev., and a good graphical extrapolation to 1·13, the ratio of which is 0·70. While the data are clearly inadequate for sweeping generalizations, we may reasonably suspect that molecular states correlated to an atom in a $^1S_0$ state tend to give an abnormally high linear Birge-Sponer extrapolation. The alkaline-earth metals mostly have low ionization potentials, and many of their compounds, especially their halides, may show partially ionic bonding which may confuse the relation between their true dissociation energies and the linearly extrapolated values.

§4. MOLECULES WITH PARTIALLY IONIC BONDING

For molecules composed of a metal with low ionization potential and a non-metal with high electron affinity the bonding force tends to be largely of an ionic nature. For all molecules except CsF the ionization potential of the metal exceeds the electron affinity of the non-metal, and the ground states are believed to dissociate to neutral atoms. Nevertheless, since the ionic forces in some of these cases are believed to be relatively important, it is probable that the potential energy curve will assume a different form from that for molecules in which ionic bonding is unimportant. Thus, it seems likely that the relation between the

true value of the dissociation energy and the linearly extrapolated value may be abnormal. In making the comparison we are again handicapped by lack of data. The most ionic molecules are the halides of the alkali metals. Good values of the dissociation energy are available for most of them, thermochemical values agreeing closely with those obtained by observation of atomic fluorescence. These molecules do not, unfortunately, readily give discrete band spectra from which values of ω_e and $x_e\omega_e$ can be obtained ; for some of them, however, Beutler and Levi (1934) have succeeded in observing some rather diffuse bands from which approximate values of ω_e and $x_e\omega_e$ have been derived. The only other molecules of this type for which data are satisfactory are TlCl, TlBr, and GaI ; the measurements are most precise in the case of TlCl, for which the dissociation energy has been derived both thermochemically and by atomic fluorescence, while the values of ω_e and $x_e\omega_e$ have been obtained by Howell and Coulson (1938) from origin measurements.

The values of D_0, D_{lin}, D_0/D_{lin}, and also the difference between the metallic ionization potential (I) and the electron affinity ($E.A.$) of the non-metal are shown in table 3. It will be seen that for most of the molecules listed the true dissociation

Table 3. Ground states of molecules with predominantly ionic bonding

Molecule	D_0	D_{lin}	D_0/D_{lin}	$I-E.A.$
LiI	3·6	4·2	0·86	1·6
NaCl	4·24	4·4	0·96	1·4
NaBr	3·83	2·6$_6$	1·43	1·5
NaI	3·11	3·4	0·91	2·1
KCl	4·40	2·7	1·63	0·6
KBr	3·93	2·2	1·78	0·7
KI	3·31	2·0	1·65	1·1
TlCl	3·80	2·1	1·81	2·3
TlBr	3·2	2·9*	1·10	2·5
GaI	2·85	2·9	0·98	2·8

* From Howell and Coulson (1941).

energy is much *higher* than that obtained by linear extrapolation ; this is particularly true for the halides of potassium, for which the values of $x_e\omega_e$ are likely to be more accurate than for the halides of Na and Li. There is some doubt about the values of D_0 for GaCl, GaBr, InCl, InBr and InI, but here again it is quite likely that the true dissociation energy is much higher than the linearly extrapolated value. For HF, although the ionization potential of hydrogen exceeds the electron affinity of fluorine by over 9 ev., it is believed that ionic bonding is fairly important; if we accept Herzberg's (1939) value for D_0 we obtain $D_0/D_{lin}=1\cdot07$; a rather lower value of D_0 appears probable to the author, but it seems fairly certain that the ratio is not less than 1.

Thus we see that values for the dissociation energies of molecules with partially ionic bonding which have been derived from a linear Birge-Sponer extrapolation should be treated with caution. In some cases such values may be too low, instead of, as is usually thought to be the case, on the high side. It is not always easy to say definitely which molecules are subject to appreciable ionic binding forces; the rough rule, suggested by Mulliken, that ionic bonding is likely to be predominant when the metallic ionization potential exceeds the electron

affinity of the non-metal by less than 3 ev., is helpful, but there are cases when such forces are still important even when the ionization potential is a good deal higher than the electron affinity, as in HF.

§5. THE EXTRAPOLATION FOR EXCITED STATES

So far we have considered the extrapolation only for the ground states of molecules. For these, the perturbations of the potential-energy curve by those of other states are likely to cause only a general deepening of the potential minimum without any violent alteration in its general form. (The special cases where the ground state correlates with an atom is a $^1S_0$ state, and where the perturbations are due to ionic terms, have been discussed already.) For excited electronic states, however, mutual interactions are more likely to cause perturbations which will modify the shape of the potential curves in an irregular way, and deviations from a smooth Birge-Sponer extrapolation of the vibrational energy levels may be expected to be more frequent.

It should be especially noted that interaction from ionic terms is likely to be more frequent and more severe for excited states. Thus, for the halides of Al, Ga, In, and Tl, the dissociation limit to ions (e.g. to $Al^+ + Cl^-$, etc.) lies below the limit going to a normal halogen atom and a metal atom in its lowest excited state* (e.g. to $Al(^2S) + Cl(^2P)$). Many of the excited states of molecules formed from one strongly electro-positive and one strongly electro-negative atom are influenced by ionic forces, and may even dissociate to ions; in this case the Birge-Sponer extrapolation would take quite a different form, and it would be quite incorrect to assume a linear convergence of the vibrational levels, of which there would in this case be an infinite number.

There are a fair number of examples of abnormalities for excited states. The most obvious group calling for comment is that of the upper states of the main systems of the alkali-metal hydrides. For all these the spacing of the vibrational energy levels at first increases so that $x_e\omega_e$ has a fairly large negative value; instead of converging, the levels diverge. The unusual form of the potential curves for these states has been attributed by Mulliken (1936) to interaction with the potential curve arising from ionic products. Although a linear extrapolation is clearly impossible, the curve of $\Delta G(v+\frac{1}{2})$ against v appears to be smooth, rising to a flat maximum and then falling fairly sharply, so that a graphical extrapolation to the dissociation limit is not necessarily ruled out; a good value can be obtained in this way for KH, and the method can also be used for NaH. For some of the excited states of the alkali-metal molecules (K_2, Li_2, etc.) a similar, but less marked, effect is observed. The vibrational energy levels at first show a slow rate of convergence which slowly increases, but at high vibrational quantum numbers the rate of convergence increases sharply and the graphical extrapolation appears to indicate a dissociation energy for these states of only about half that given by linear extrapolation.

The most extreme anomaly is for the excited state of AgH. From the analysis by Bengtsson and Olsson (1931) it appears that the vibrational levels at first show rapid and increasing convergence, but the curve of $\Delta G(v+\frac{1}{2})$ against v flattens out

---

* By lowest excited state is meant the lowest main term other than the ground one, ignoring multiplet splitting; for Al, etc., the ground state is ^2P and the lowest excited state ^2S.

abruptly at higher v, and the rate of convergence must be practically zero for a number of levels before increasing again (see Herzberg, 1939; Gaydon, 1946).

It has been shown (Herzberg and Mundie, 1940) from the form of the limiting curve of dissociation, as obtained from predissociations, that the excited $^1\Pi$ states of AlH and BH possess potential maxima. Mulliken (1937) has shown that some of the potential curves of the excited states of the silver halides, and probably the copper halides, possess quite marked maxima. In these cases even a graphical Birge-Sponer extrapolation would be of no avail, as it would lead to the energy of the potential maximum instead of to the dissociation·limit. The author is inclined to think that the potential curves of the excited states of many other halides, including probably those of Al, B, Ga and In, may possess similar maxima.

Although it is clear that there are many instances in which the extrapolation (especially the linear form) is likely to fail when applied to excited molecular states, there are still many for which it seems to work satisfactorily. As examples may be quoted the upper states of the main systems of O_2, P_2 and the various halogen molecules; it is true that at one time Birge (1929) pointed out certain irregularities in the curves of $\Delta G(v+\frac{1}{2})$ against v for some of these, but these irregularities have been largely smoothed away by later experimental work, and were in any case not of a magnitude seriously to affect the accuracy of the extrapolation. Many of the lower excited states of N_2 and CO give satisfactory extrapolations to the expected limits when the author's (Gaydon and Penney, 1945; Gaydon, 1946) latest values for the dissociation energy are used.

In general it seems that where a reasonable extrapolation from ground state data is available it is to be preferred to one for an excited state, but for molecules with atomic (non-ionic) binding, especially homonuclear molecules and molecules formed from non-metallic elements, extrapolations from the lower excited states are permissible. For metallic halides, and probably hydrides and oxides as well, it is unsafe to place reliance on extrapolations for excited states.

§6. THE EXTRAPOLATION FOR IONIZED MOLECULES

So far we have considered the extrapolations for neutral molecules which dissociate to two neutral atoms. For singly ionized molecules, however, dissociation will take place to one neutral atom and one charged atom. The force between a charged and a neutral atom at large distances is likely to be much greater than

Table 4. Ionized molecules

Molecule	D_0	D_{lin}	D_0/D_{lin}
H_2^+	2·649	2·50	1·06
HCl^+	4·5	4·0	1·12
O_2^+	6·48	6·48	1·00
N_2^+	8·73	9·2	0·95
CO^+	8·2	9·8	0·84

that between two neutral atoms. Thus the potential-energy curve at large internuclear distances is likely to take a rather different form for ionized molecules, and it would seem best to make a separate comparison of the true and linearly extrapolated values of the dissociation energy for this type of molecule. The few available values are collected in table 4.

For O_2^+ the value of D_0 is derived from a good graphical extrapolation for the excited state; it is probably of quite high accuracy, but does, of course, assume a normal trend for the extrapolation. The values for N_2^+ and CO^+ are derived from the values of $D(N_2)$ and $D(CO)$ given by Gaydon and Penney (1945) combined with the ionization potentials of N_2 and CO; for CO the value 14·1 ev. for $I(CO)$ has been used, but it is just possible that the value is 14·55 ev.

From the values collected in table 4 there appears to be little evidence for any systematic deviation from the linear value. It would therefore seem that ionized molecules are rather more suitable for making a linear extrapolation than neutral molecules, for which the true value, as we have seen, tends to be around 20 % below that of the linear extrapolation. However, all the molecules included in table 4 are relatively light. For BeH^+, MgH^+ and CdH^+ accurate values of D_0 are not available, but the rate of convergence of the ground state vibrational energy levels does increase slightly with v, and the linear extrapolation is probably a little high. Most probably the linear extrapolation is fairly good for light ionized molecules and a little high for the heavier ones, although not to the same extent as for the corresponding neutral molecules.

§7. THE RELATION BETWEEN THE FORM OF THE EXTRAPOLATION AND THE POTENTIAL-ENERGY CURVE

Clearly, the rate of convergence of the vibrational energy levels must be related to the shape of the potential-energy curve for the particular molecular state. Now for the lower vibrational levels the rate of convergence is frequently nearly a linear function of the vibrational quantum number, and we have been chiefly concerned with the linear extrapolation. One of the simplest functions for representing the potential-energy curve of a diatomic molecule is that due to Morse, which is of the form

$$U = D_e \, (1 - e^{-\beta(r-r_e)})^2.$$

One of the main virtues of this function is that when inserted in the wave equation it leads to a series of vibrational energy levels which do converge linearly. Thus it seems desirable to make use of this simple function as a basis for comparing the form of the potential curve and departures from linear convergence of levels.

For normal molecules, in which ionic binding forces are unimportant, the lower vibrational levels usually show a very nearly linear rate of convergence, but, as we have seen, extrapolation of this convergence leads to a dissociation energy of the order 20 % too high. We may take the ground state of O_2 as a typical example. For this, some 21 vibrational levels are known up to over 3·4 ev., and over this range the rate of convergence is nearly linear, showing only a slight sign of an increased rate at the highest observed values of v. Figure 1 shows a curve (dashed line) obtained from a Morse function (which uses the values of D_e, ω_e and r_e, which are derived from the three spectroscopic constants D_0, ω_e and B_e) and a curve (full line) plotted from a modified function proposed by Hulburt and Hirschfelder (1941) which uses the five spectroscopic constants D_0, ω_e, $x_e\omega_e$, B_e and α_e; the five-constant function can be relied upon in this case to come very close to the true potential curve. It will be seen from figure 1 that at $r > r_e$ the true potential curve lies above that given by the Morse function. The true curve tends to approach the dissociation limit more rapidly than the Morse

A. G. Gaydon

curve, and this more rapid approach to the limit is associated with a less rapid rate of convergence of the lower vibrational levels and an increasing rate of convergence of the higher ones.

For molecules dissociating to atoms of which one is in a $^1S_0$ state, the vibrational
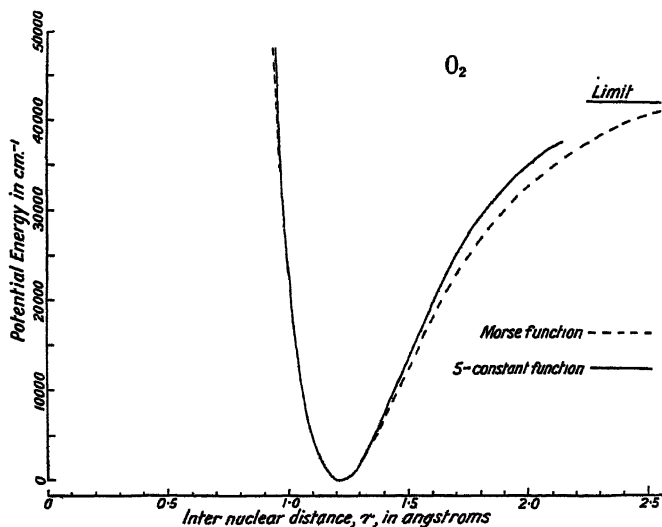
Figure 1. Potential-energy curves for the ground state O_2.

levels converge with unusual rapidity; even for the lower levels there is usually a marked departure from a linear rate of convergence, and the linear extrapolation, for which there is really little justification, comes much too high. Clearly, in these cases the potential curves must depart considerably from the Morse form.
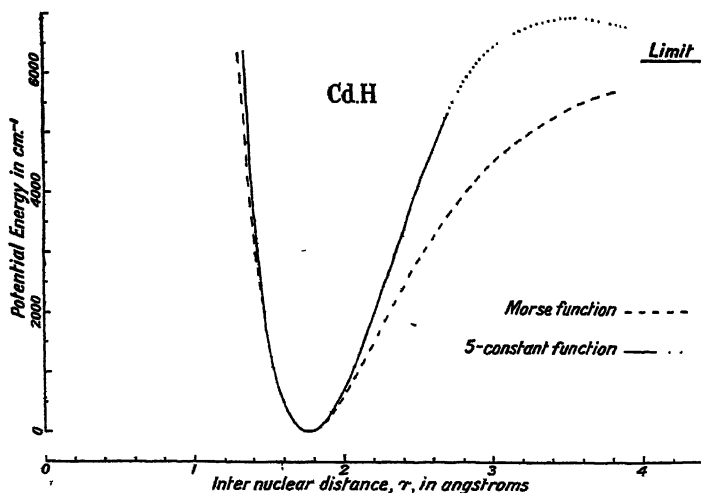
Figure 2. Potential curves for the ground state of CdH.

Owing to this big difference it is not easy to plot the potential curve, even when experimental data are adequate, as for many of the alkaline-earth hydrides. Figure 2 shows the potential curve for the ground state of CdH drawn to a Morse function (dashed line) and also a curve drawn to a five-constant Hulburt-

Hirschfelder function (full line, with upper part dotted). This latter curve probably approaches the true curve quite well around the minimum, but may be subject to appreciable error as r increases, and the potential maximum to which it leads should not be taken seriously. However, it serves to show that the strong departure from a linear rate of convergence of the vibrational levels is again associated with a much more rapid rise of the potential curve towards the dissociation limit than that given by the Morse function.

The differences between the actual potential curve and the Morse form so far considered have been in the region where r is greater than r_e but is still not very large ($\sim 2r_e$); this is probably the region of most importance for the extrapolation, and here the tendency is for the rate of convergence to be initially less than that given by the Morse function, but to increase and exceed it at high v, and for the true potential-energy curve to lie above that of the Morse function. At larger internuclear distances there may be a tendency for the reverse effect. The Morse function corresponds to an exponential falling off of the potential energy at very large r. The actual potential energy at very large distances is, however, likely, ·on account of interaction of two atoms, to fall off much less rapidly than that given by the exponential law. The weak van der Waals forces between atoms are known (see, for example, Margenau and Watson, 1936) to fall off as $1/r^7$, corresponding to a potential energy varying as $1/r^6$. There may be other forces, which vary as a still lower power of $(1/r)$: thus the resonance interaction between like atoms in different states gives a force proportional to $1/r^4$; polarization forces also are likely to be important in many cases, and these will again give a potential energy which, at large distances, falls off less rapidly than it would if it obeyed the exponential law. The force between two atoms in P states may vary as $1/r^6$. Thus the true potential curve should cross the Morse curve somewhere at large internuclear distances and then approach the dissociation limit less rapidly. It would not, therefore, be surprising if towards the limit the rate of convergence of the vibrational levels were to decrease slightly.

There are not many examples for which we have adequate data on the vibrational levels very near to the dissociation limit, but there is some evidence to support this conclusion. For the upper state of O_2 and for the ground state of H_2 Birge (1929) assumed an inflection in the curve of $\Delta G(v+\frac{1}{2})$ against v at high v, but later experimental work has not provided much support for any marked change, although a slight trend towards a slower rate of convergence may exist. For the upper states of the alkali metals (K_2, Li_2) there is no evidence for a decreased rate of convergence—indeed, the reverse is true. For the upper states of the halogens and mixed halogen molecules, however, there is definite evidence of a decrease in the rate of convergence of the vibrational levels as the limit is approached. This occurs for the upper states of the main systems of all the halogen molecules for which data exist.

The effect is well illustrated by ICl; the curve of $\Delta G(v+\frac{1}{2})$ against v for the upper state is shown in figure 3, in which data are from Darbyshire (1932). The graph at first shows negative curvature corresponding to a slow but steady increase in the rate of convergence of the levels; between about $v=20$ and $v=25$ the graph shows an inflection and the " toe " turns up so that the rate of convergence (i.e. the slope) decreases to a value below that at $v=0$. This tendency for a more

positive curvature to the graph in the region of the " toe " would tend, if it occurred for other molecules, to bring the dissociation energy a little higher than that which would be obtained by a graphical Birge-Sponer extrapolation using only points before the " toe " was reached. The effect must, however, be very small. The laws of force which we have discussed will only apply at large values of r where the potential energy is already small, and the effect on the dissociation energy will be almost negligible. Even in the case of ICl, which is one of the most extreme examples, the " toe " represents only about 0·04 ev., and the difference in dissociation energy obtained by using the true curve and by using an extrapolation from before the inflection (i.e. from around $v = 20$) would amount to only about 0·02 ev. There is little reason to think that this effect is important for any neutral molecule.

As we have already pointed out, ionized molecules dissociate to one charged atom and one neutral atom. The force between these atoms at large distances is likely to be that between a charge and an induced dipole, which is
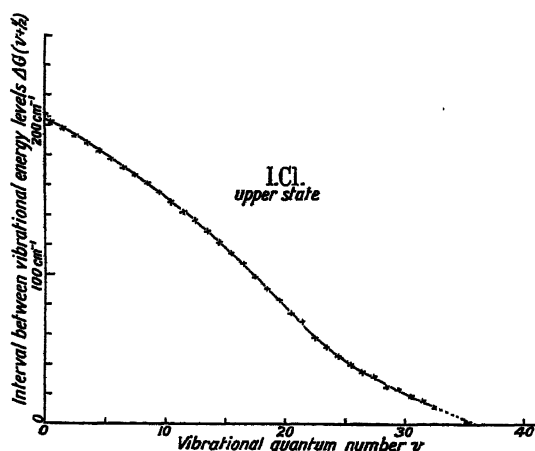
Figure 3. Graph of $\triangle G(v+\frac{1}{2})$ against v for the upper state of ICl.

proportional to $1/r^4$. It is, therefore, especially likely that for ionized molecules the curve of $\Delta G(v+\frac{1}{2})$ against v will show an inflection towards a slightly higher limit. There are no available data to check this supposition, but the fact that for the lighter ionized molecules the linear extrapolation does not show any systematic tendency to be too high may be connected with the existence of "toes" to the curves. If this is so, the dissociation energy of O_2^+, which depends on a good graphical extrapolation for the excited state, may be a little higher than is generally assumed, although any considerable change would lead to a discrepancy between the spectroscopic and the electron-impact values for the ionization potential of O_2.

For molecules in which ionic bonding is important there is no case in which the data are adequate for the drawing of an accurate potential curve. In many cases, however, the low value of the linear extrapolation seems likely to be associated with the fact that the true potential curve lies below the Morse function curve in the region where r is a little greater than r_e.

§8. SUMMARY

Although the remarks contained here may appear to be a severe criticism of the Birge-Sponer extrapolation, they should really lead to an increase in its usefulness

by giving a fuller understanding of the limitations of the method and the likely deviations from the value given by linear extrapolation.

Where data are adequate for a good graphical extrapolation for a ground state, the result should be satisfactory. Where, for a non-ionic molecule, the lower levels appear to converge linearly, the true dissociation energy tends to be around 80 % of that given by linear extrapolation, being perhaps a little nearer this extrapolated value for light molecules and rather lower for heavy molecules. When dissociation leads to an atom in a $^1S_0$ state the levels do not usually converge linearly; the true dissociation energy is then often not much over half the linearly extrapolated value. Extrapolations for excited states should be treated with caution, especially when ionic forces are likely to be important; satisfactory values may be obtained for the lower excited states of molecules composed of non-metals, especially if a graphical method is used .

For molecules with largely ionic bonding the linear Birge-Sponer extrapolation does not appear satisfactory; in a number of examples it gives a value which is too low. For ionized molecules there appears little evidence to suggest any failure of the linear extrapolation.

For most types of molecule the above considerations should assist in the estimation of the dissociation energy, taking the linearly extrapolated value as a basis, and should also help in the assessment of the reliability of the value.

There are two groups of molecules for which data are still inadequate to give much guidance. For metallic oxides there will usually be some tendency towards ionic bonding. The electron affinity of atomic oxygen is not quite certain, but definitely exceeds 2 ev. and may be nearer 3; the ionic bonding forces are probably not very important, provided the metallic ionization potential exceeds about 7 ev., and the linear extrapolation can then be expected to be a little too high, as usual; for the oxides of the more electropositive metals any figure based on linear extrapolation should be taken as only provisional. The halides of the elements of Group II are another unsolved problem, as we have the complexities due both to ionic bonding and also to the presence of an atom with $^1S_0$ ground state; in many of them, too, the value of $x_e \omega_e$ derived from head measurements may be inaccurate owing to the systems being weakly degraded. The evidence for CaCl and CaF suggests that the linear extrapolation may be appreciably too high, as in the hydrides of Group II elements, but this may not always be so.

ACKNOWLEDGMENTS

REFERENCES

BENGTSSON, E. and OLSSON, E., 1931. *Z. Phys.*, **72**, 163.
BEUTLER, H. and LEVI, H., 1934. *Z. Phys. Chem.*, B, **24**, 263.
BIRGE, R. T., 1929. *Trans. Faraday Soc.*, **25**, 707.
BIRGE, R. T. and SPONER, H., 1926. *Phys. Rev.*, **28**, 259.
DARBYSHIRE, O., 1932. *Phys. Rev.*, **40**, 366.
GAYDON, A. G., 1946. *Dissociation Energies* (London: Chapman and Hall) (in press).
GAYDON, A. G. and PENNEY, W. G., 1945. *Proc. Roy. Soc.*, A., **183**, 374.

HERZBERG, G., 1939. Molecular Spectra and Molecular Structure (New York: Prentice-Hall).
HERZBERG, G. and MUNDIE, L. G., 1940. J. Chem. Phys., 8, 263.
HOWELL, H. G. and COULSON, N., 1938. Proc. Roy. Soc., A., 166, 238.
HOWELL, H. G. and COULSON, N., 1941. Proc. Phys. Soc., 53, 706.
HULBURT, H. M. and HIRSCHFELDER, J. O., 1941. J. Chem. Phys., 9, 61.
KRATZER, A., 1924. Z. Phys., 26, 40.
MARGENAU, H. and WATSON, W. W., 1936. Rev. Mod. Phys., 8, 22.
MULLIKEN, R. S., 1936. Phys. Rev., 50, 1017 and 1028.
MULLIKEN, R. S., 1937. Phys. Rev., 51, 310.
RYDBERG, R., 1931. Z. Phys., 73, 376.
WAHRHAFTIG, A. L., 1942. J. Chem. Phys., 10, 248.
WIELAND, K., 1941. Helv. Phys. Acta, 14, 420.

ULTRA-VIOLET ABSORPTION BAND-SYSTEMS OF SiS, SiSe AND SiTe

By E. E. VAGO and R. F. BARROW,

Physical Chemistry Laboratory, University of Oxford

ABSTRACT. A new ultra-violet band-system, E–X, of each of the molecules SiS, SiSe and SiTe has been observed in absorption at 800–1000° C., and vibrational analyses have been done. The D–X systems, already known in emission, have also been photographed in absorption. These observations have extended the D–X systems of SiSe and of SiTe and have led to somewhat improved values of the vibrational constants. The results are summarized in the following table :—

Molecule	State	ν_e	ω_e	$x_e\omega_e$	$y_e\omega_e$	D_e
SiS	E	42322·1	402·3	−1·74	−0·025	1·2 ev.
	D $^1\Pi$	35028·8	512·0	−2·38	−0·045	2·3
	X $^1\Sigma$	0	749·5	−2·56	—	6·6
SiSe	E	38505·9	308·8	−1·95	−0·032	1·3
	D	32450·3	399·8	−1·93	—	2·0
	X ($^1\Sigma$)	0	580·0	−1·78	—	6·0
SiTe	E	33991	242	(−3·63)	(+0·13)	(1·4)
	D	28661·8	338·6	−1·70	—	2·0
	X ($^1\Sigma$)	0	481·2	−1·30	—	5·6

§1. INTRODUCTION

SiS, SiSe and SiTe each give rise to one band-system (D–X) in the near ultra-violet region, which appears readily in emission in positive-column discharges. For several reasons, including the fact that two near ultra-violet absorption systems of GeS (which is isosteric with SiSe) were known, we thought it likely that the absorption spectra of the silicon compounds would contain systems lying at shorter wave-lengths than the D–X systems already known. This has proved so : SiS, SiSe and SiTe each show banded absorption in the ultra-violet region, and the bands of each molecule may be assigned to two systems D–X and E–X.

The D–X system of SiS was originally developed rather well in emission, and the absorption measurements have added nothing to the vibrational analysis of Barrow and Jevons (1938). Development of the D–X systems of SiSe and of SiTe

in emission was, however, somewhat incomplete (Barrow, 1939), and the absorption work has supplemented the original vibrational analyses and has led to improved values of the constants.

Vibrational analyses of the three new E–X systems have been carried out, and a preliminary account of the work has been published (Vago and Barrow, 1946). The detailed results are given and discussed in the present paper.

§2. EXPERIMENTAL METHODS AND RESULTS

The absorption spectra were studied at temperatures in the range 800–1000° c. by methods already described (Barrow and Vago, 1944). An account of the preparation of SiS has been given elsewhere (Barrow and Jevons, 1938). SiSe and SiTe were made via the magnesium compounds. A mixture of finely-divided Mg and Se or Te was heated gently, using silica as a diluent to moderate the violence of the reaction. The products were mainly MgSe or MgTe mixed with silica: they were heated *in vacuo* at 1100° c. for about three hours to form the silicon compounds. These were found to sublime only very slowly, and since the other components of the reaction mixture were not volatile, the crude products were used without purification. SiSe and SiTe proved objectionable substances, as, like SiS, they hydrolyze rapidly in moist air.

A hydrogen tube was used as source of continuous radiation. The spectra were photographed in (i) a Hilger medium quartz prism instrument (E. 498), (ii) a first order of a concave grating spectrograph with a linear dispersion of about 7·4 a./mm., and (iii) in a large quartz Littrow spectrograph with a dispersion of about 2·4 a./mm. at 2500 a. Ilford Ordinary, Process and Q.1 plates were used: they were developed in Process developer to secure maximum contrast.

The absorption spectra of all three molecules showed two distinct band-systems, but, whereas not much difficulty was experienced in getting good spectrograms of the three D–X systems and of the E–X system of SiS, strong contrast was never obtained on plates of the E–X system of SiSe; and the E–X system of SiTe was even harder to photograph. These difficulties seemed to be due to the intrinsic weakness of the systems: they could probably have been overcome by the use of longer columns of vapour, but in fact the spectrograms of the new systems of SiSe and SiTe proved adequate for their vibrational analyses. A reproduction of the E–X system of SiS is given in plate 1: the other E–X systems are generally similar in appearance.

The measurements of the E–X systems are set out in the Delandres schemes in tables 1–3. Additional or amended measurements of band-heads of the D–X systems of SiSe and SiTe are given in table 4: these confirm the original v', v'' assignments. There can be few doubts about the correctness of the analyses given for the E–X systems, since the ground-state differences are also known from the D–X systems. Unfortunately, however, the distribution of intensities in the E–X systems and the populations in the initial levels (determined by the Boltzmann law) combine to make the $v'=0$ progressions very difficult to observe. Indeed, we failed to measure any band-heads of this progression on the first plates of the SiS system, and they were only found on later plates of more contrast of the absorption at somewhat higher temperatures. It is therefore possible, but, we think, unlikely, that we have failed to observe $v'=0$ progressions in the E–X systems of SiSe and SiTe.

Table 1. Measurements of band-heads of the
E←X system of SiS

Italic numerals denote wave-lengths in air (I.A.);
Large Roman, wave-numbers *in vacuo* (cm.⁻¹);
Small Roman, wave-number differences.

v'	0		1		2		3
15	47664 *2097·35*						
	326						
14	47337·9 *2111·80*						
	336·9						
13	47001·0 *2126·94*						
	343·3						
12	46657·7 *2142·59*						
	349·4						
11	46308·3 *2158·76*						
	355·5						
10	45952·8 *2175·46*						
	360·6						
9	45592·2 *2192·67*						
	365·2						
8	45227·0 *2210·38*						
	369·4						
7	44857·6 *2228·58*	744·5	44113·1 *2266·20*				
	374·8		375·6				
6	44482·8 *2247·36*	745·3	43737·5 *2285·66*	740·2	42997·3 *2325·01*		
	378·7		376·3		375·5		
5	44104·1 *2266·66*	742·9	43361·2 *2305·50*	739·4	42621·8 *2345·50*		
	384·6		384·2		385·0		
4	43719·5 *2286·60*	742·5	42977·0 *2326·11*	740·2	42236·8 *2366·88*	735·9	41500·9 *2408·85*
	384·8		386·8		387·2		385·8
3	43334·7 *2306·91*	744·5	42590·2 *2347·24*	740·6	41849·6 *2388·78*	734·5	41115·1 *2431·46*
			391·4		390·8		391·6
2			42198·8 *2369·01*	740·0	41458·8 *2411·30*	735·3	40723·5 *2454·84*
			394·4		394·7		393
1			41804·4 *2391·36*	740·3	41064·1 *2434·48*	733	40331 *2478·7*
					395		399
0					40669 *2458·1*	737	39932 *2503·5*

$v' \uparrow$

$v'' \to 0$ 1   2   3

Table 2. Measurements of band-heads of the E ← X system of SiSe

* The 3,0 and 5,1 bands are coincident.

v'	v''=0		v''=1		v''=2		v''=3		v''=4
8	40698·0 / 2456·38								
	276·8								
7	40421·2 / 2473·20								
	284·8								
6	40136·4 / 2490·75								
	287·7								
5	39848·7 / 2508·74	575·2	39273·5* / 2545·48	570·7	38702·8 / 2583·02				
	286·1		285·6		287·5				
4	39562·6 / 2526·88	574·7	38987·9 / 2564·13	572·6	38415·3 / 2602·35	568·2	37847·1 / 2641·42	568·1	37279·0 / 2681·68
	289·1		290·8		290·9		290·8		
3	39273·5* / 2545·48	576·4	38697·1 / 2583·40	572·7	38124·4 / 2622·21	568·1	37556·3 / 2661·88		
	295·5		297·7		298·7		295·6		
2	38978·0 / 2564·78	578·6	38399·4 / 2603·43	573·7	37825·7 / 2642·92	565·0	37260·7 / 2683·00		
			301·4		299·0		302·0		
1	—		38098·0 / 2624·03	571·3	37526·7 / 2663·98	568·0	36958·7 / 2704·92		
			305·3		305·3		305·9		
0	—		37792·7 / 2645·23	571·3	37221·4 / 2685·83	568·6	36652·8 / 2727·50	556·5	36096·3 / 2769·65

Table 3. Measurements of band-heads of the E ← X system of SiTe

v'	v''=0		v''=1		v''=2		v''=3		v''=4
6	35210·1 / 2839·26	478·0	34732·1 / 2878·34						
	218·2		218·2						
5	34991·9 / 2856·97	478·0	34513·9 / 2896·54						
	219·5		218·2						
4	34772·4 / 2876·00	476·7	34295·7 / 2914·97						
	219·7		221·0						
3	34552·7 / 2893·28	478·0	34074·7 / 2933·87						
	222·5		220·2						
2	34330·2 / 2912·04	475·7	33854·5 / 2952·96	475·1	33379·4 / 2994·99				
			225·5		226·8				
1			33629·0 / 2972·76	476·4	33152·6 / 3016·48	472·5	32680·1 / 3059·08	470·0	32210·0 / 3103·73
			233·2		234·1		236·9		
0			33395·8 / 2993·52	477·3	32918·5 / 3036·92	475·3	32443·2 / 3081·42		

The electronic and vibrational constants for the three known states of these molecules are collected together in table 5. The new data on the D–X and E–X systems of SiSe and of SiTe have led to slight changes in the values of the constants for the D and X states of these molecules. The ground-state constants for SiS determined from the extensive D–X systems were found to be in good agreement with the figures found from the analysis of the E–X system. The E-state data for

Table 4. Additional or amended measurements of band-heads of the
D–X systems of SiSe and SiTe

	SiSe			SiTe	
v', v''	v		v', v''	v	
8,0	35431·6 cm.$^{-1}$		7,0	30867·8 cm.$^{-1}$	
7,0	055·1		8,1	700·1	
8,1	34854·0		6,0	549·3	
6,0	679·1		7,1	385·0*	
7,1	476·0		5,0	232·3*	
5,0	302·2*		6,1	070·8*	
6,1	101·1*		4,0	29910·4*	
5,1	33727·1		5,1	752·4	
4,1	346·1		3,0	586·3*	
			4,1	431·6	
			2,0	257·4*	
			3,1	106·9	

* Band-head also measured in emission.

Table 5. Summary of electronic and vibrational constants

Molecule	State	v_e	ω_e	$x_e\omega_e$	$y_e\omega_e$	κ_e
SiS	E	42322·1	402·3	−1·74	−0·025	$1·42 \times 10^5$
	D	35028·8	512·0	−2·38	−0·045	2·31
	X	0	749·5	−2·56	—	4·94
SiSe	E	38505·9	308·8	−1·95	−0·032	1·16
	D	32450·3	399·8	−1·93	—	1·96
	X	0	580·0	−1·78	—	4·11
SiTe	E	33991	242	(−3·63)	(+0·13)	0·79
	D	28661·8	338·6	−1·70	—	1·55
	X	0	481·2	−1·30	—	3·13

Note.—The units are cm.$^{-1}$ (*in vacuo*), except for κ_e, the force constant, which is given in dynes/cm. The signs of the vibrational constants are those applicable when the vibrational energy levels are represented by

$$[\omega_e(v+\tfrac{1}{2})+x_e\omega_e(v+\tfrac{1}{2})^2+y_e\omega_e(v+\tfrac{1}{2})^3+ \ldots].$$

SiS and SiSe call for no particular remark. The expression given for the vibrational energy in the E state of SiTe is, however, inherently rather unlikely, as it includes unexpectedly large anharmonic and cubic terms. Errors in these terms might be attributed to unsuspected perturbations, or to the incorrect location of some of the band-heads (the Te isotope effect might be partly responsible). However, the expression given represents satisfactorily the experimental measurements, and at the worst the value of ω_e is not likely to be much in error.

§3. DISCUSSION

It was largely the existence of two absorption systems of GeS which led us to look for second absorption systems of the present molecules: in particular, the expectation was that SiSe, being isosteric with GeS (48 electrons), would be spectroscopically rather similar. How close the constants are for these

compounds may be judged from the following figures :—

State	GeS		SiSe	
	ν_e	κ_e	ν_e	κ_e
E	4·82 ev.	1·27 × 10⁵	4·77	1·16 × 10⁵ dyne/cm.
D	4·08	1·85	4·02	1·96
X	0	4·36	0	4·11

The isosteric group SiTe, GeSe and SnS show analogous behaviour in regard to their electronic and vibrational constants. The rotational constants of such groups of molecules may also be expected to lie rather close together.

Lacking rotational analyses, we can only speculate about the molecular types of the E states of these molecules. One possibility—and it is no more than that—may, however, be suggested. The lowest known excited states of SiO and of SiS are $^1\Pi$, and the presumption is therefore strong that these states (and probably the similar states of SiSe and SiTe) correspond to the lowest excited singlet state of CO, A$^1\Pi$. Now, although the bond in unexcited CO is much stronger than, for example, that in unexcited SiS, it might perhaps be expected that the *relative* change in bond strength in going from the unexcited $^1\Sigma$ level to the $^1\Pi$ level should be similar for the two molecules. To test this possibility we set the force constants for the ground states at 100, and calculate those for the excited states as percentages of the ground-state force constants. The figures are given in table 6.

Table 6. Force constants, relative to those for the ground states

CO		SiO		SiS		SiSe		SiTe	
A$^1\Pi$	48·8	$^1\Pi$	47·0	D$^1\Pi$	46·8	D	47·7	D	49·5
$a'\,^3\Sigma$	29·7	—		E	28·8	E	28·2	E	25
$a\,^3\Pi$	64·3	—		—		—		—	
x$^1\Sigma^+$	100	$^1\Sigma$	100	x$^1\Sigma$	100	x	100	x	100

It appears that for the three molecules CO, SiO and SiS the force constant for the $^1\Pi$ level is about 48% of that in the ground state: the values for the D levels of SiSe and SiTe suspected to be $^1\Pi$ are also close to 48%. The figures for the E levels are much lower and agree fairly well with that for the $a'\,^3\Sigma$ state of CO. This measure of agreement may be quite fortuitous—it is noteworthy that the $a'\,^3\Sigma$ level lies somewhat *below* the A$^1\Pi$ level in CO, whereas the E states for the silicon compounds lie a little *above* the corresponding D states; but for what it is worth it suggests that the E states of SiS, SiSe and SiTe are of the type $^3\Sigma$, analogous to the $a'\,^3\Sigma$ state of CO.

The energies of dissociation of the silicon compounds are of interest. We have, it is true, no very precise information, but only figures derived from long extrapolations of the vibrational levels: nevertheless some tentative conclusions can be reached. Let us consider SiS (see also Barrow and Jevons, 1938). The energy of dissociation for the ground state, D_X, calculated by the Birge-Sponer extrapolation, is 6·72 ev. The values for the D and E states, calculated in the same way, but including the cubic terms, are 1·75 and 1·36, giving $(\nu_e + D)_D = 6·09$ and $(\nu_e + D)_E = 6·61$. Now it may be presumed that the ground state of SiS is formed from Si and S atoms in their ground states, ^3P. It is also most probable that the D$^1\Pi$ state dissociates into two ^3P atoms, since the next combination of atomic states that could give a singlet molecular state is, by the Wigner-Witmer rules, $^1D + ^1D$, which lies at about 2 ev. above $^3P + ^3P$. If the combination $^1D + ^1D$ were involved, we should have, using $D_X = 6·7$, $(\nu_e + D)_D = 8·7$ or $D_D = 4·4$,

which seems hopelessly out of agreement with the extrapolated value of 1·75. Thus it is most likely that the D $^1\Pi$ state of SiS also dissociates into $Si(^3P) + S(^3P)$.

The evidence about the E state is much less conclusive. It may be a triplet state, in which case it could arise either from $^3P + ^3P$ or from $^3P + ^1D$ atomic states. However, the figure for $(\nu_e + D)_E$, 6·6 ev. is close to the mean of $(\nu_e + D)_D$ and D_X, 6·4, so that the dissociation $SiS(E) \rightarrow Si(^3P) + S(^3P)$ is feasible, and the data for SiSe (which molecule may be presumed to behave similarly to SiS) suggest that this is the most reasonable hypothesis.

The extrapolated values of D for SiO, SiS, SiSe and SiTe are given in table 7.

Table 7. Energies of dissociation (ev.)

Molecule : SiO   SiS

State	$D_{\text{extrapolated}}$	$\nu_e + D$	D_{mean}	$D_{\text{extrapolated}}$	$\nu_e + D$	D_{mean}
E	—	—	(1·7)	1·36	6·61	1·22
D	—	—	2·95	1·75	6·09	2·13
X	8·26	8·26	8·26	6·72	6·72	6·47
					6·47	

Molecule : SiSe   SiTe

State	$D_{\text{extrapolated}}$	$\nu_e + D$	D_{mean}	$D_{\text{extrapolated}}$	$\nu_e + D$	D_{mean}
E	0·89	5·67	1·27	—	—	1·37
D	2·57	6·59	2·02	2·09	5·64	2·03
X	5·86	5·86	6·04	5·52	5·52	5·58
			6·04			5·58

On the basis of the arguments outlined above, we have assumed that each of the X, D and E states of all these molecules gives a pair of unexcited atoms on dissociation, and have then determined mean values of D_X. These mean values have been used to calculate the improved values of D_D and D_E given in the columns "D_{mean}". For the SiO(D) and SiTe(E) states, large cubic terms make extrapolation impossible: the figure given for SiO(E) comes from a rough prediction of that level (see below). There may be considerable errors in these figures, but they probably indicate correctly the general trends. Comparison with the force constants given in table 5 tends to show—reasonably enough—that in a given state the most tightly-bound molecule in the equilibrium position often requires also the greatest amount of energy to dissociate it into isolated atoms. Similarly, for any single molecule, the X state is more stable in both senses than the D state, which is in turn more stable than the E state.

Finally, it would seem to be fairly safe to predict the existence of an E–X system of SiO at shorter wave-lengths than the $^1\Pi - ^1\Sigma$ system. Crude extrapolation puts the E level at about 6·6 ev. above the ground state, so that the 0,0 band of this system should lie at about 1850 A. The vibration frequency in the E state should be about 670 cm.$^{-1}$ and the energy of dissociation *ca.* 1–2 ev.

REFERENCES

BARROW, R. F. and JEVONS, W., 1938. *Proc. Roy. Soc.*, A, 169, 60.
BARROW, R. F., 1939. *Proc. Phys. Soc.*, 51, 267.
BARROW, R. F., and VAGO, E. E., 1944. *Proc. Phys. Soc.*, 56, 79.
VAGO, E. E. and BARROW, R. F., 1946. *Nature, Lond.*, 157, 77.
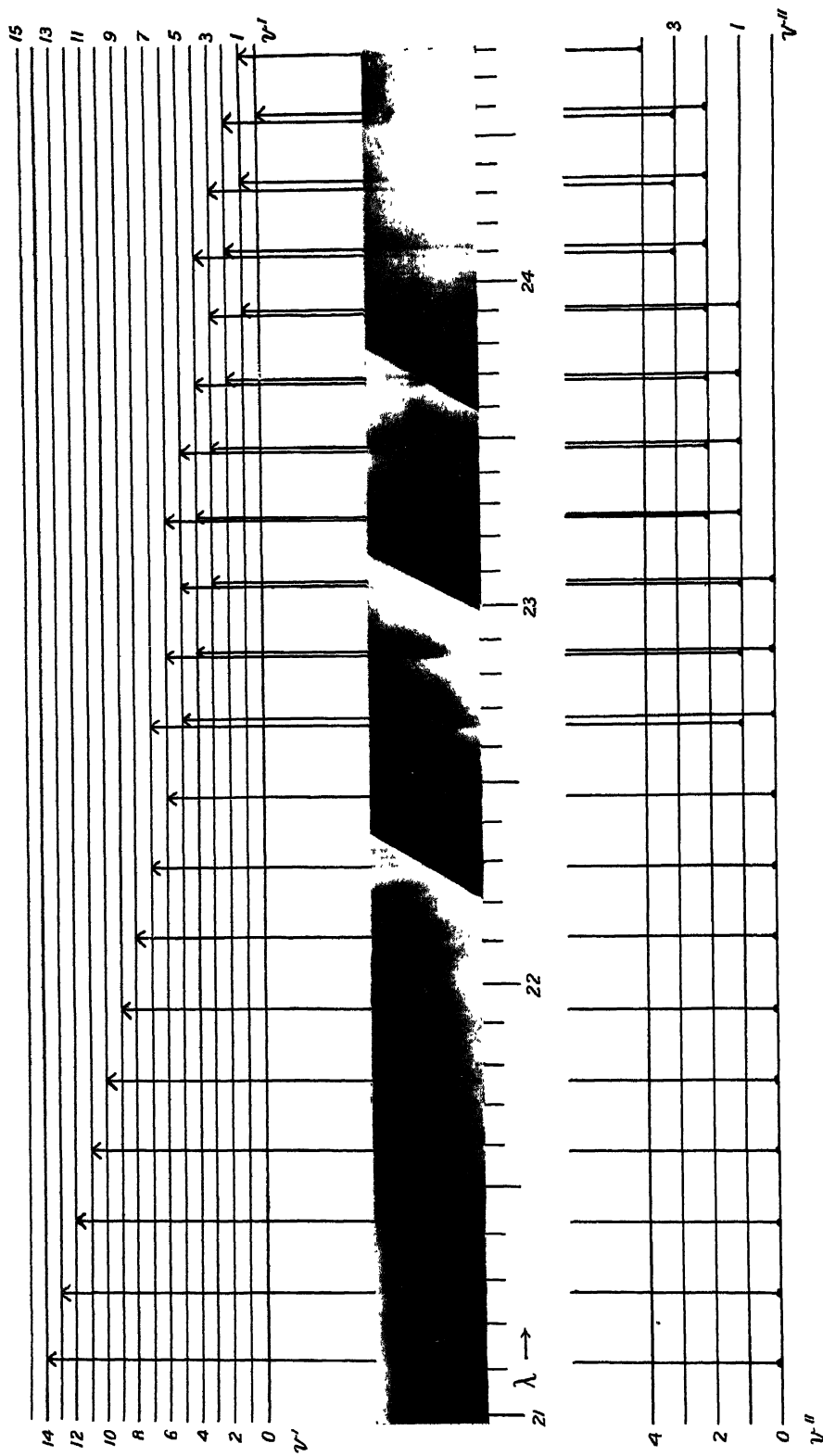
Bands of the E ← x system of SiS in absorption (medium quartz prism spectrograph).

THE ALGEBRAIC ANALYSIS OF THE HIGHER-ORDER ABERRATIONS OF OPTICAL SYSTEMS

TANGENTIAL ABERRATIONS OF A SYSTEM OF COAXIAL SPHERICAL REFRACTING SURFACES

By H. A. BUCHDAHL,

Ministry of Munitions Annexe 9/101;
Physics Department, University of Tasmania

M.S. received 16 November 1945

ABSTRACT. The behaviour of an optical system is known if the position of the point of intersection of an arbitrary ray passing through the system with any chosen plane of reference can be determined. In the present paper tangential aberrations of any order are defined in terms of the distance of this point from an ideal image point, using elementary methods only. It is shown how these aberrations may be computed in practice without the use of any trigonometrical traces.

As a practical example, the computations necessary for the determination of the complete primary and secondary aberrations and of the tertiary spherical aberration are given in detail for the case of a Cooke Triplet. The information so obtained is compared with that yielded by strict trigonometrical tracing, illustrating the good agreement between them and the simplicity of the computations involved.

---

§1. INDEX OF SYMBOLS

THE following list gives most of the symbols employed in the text. Each number in brackets refers to the paragraph in which the symbol first occurs.

Symbol	Meaning of symbol
a_ν	coefficient of $y^{3-\nu}u^\nu$ in the expansion of v'. (11)
a (suffix)	extreme ray of pencil. (3)
α	a quantity linearly related to y and u. (4)
A	$\sum\limits_{\nu=1}^{3} a_\nu p_1^{3-\nu}$. (12)
A_n	$\dfrac{1}{n!}\cdot\dfrac{d^n A}{dp_1^n}$. (13)
b_ν	coefficient of $y^{5-\nu}u^\nu$ in the expansion of v'. (11)
b (suffix)	extreme ray of pencil. (3)
B	$\sum\limits_{\nu=1}^{5} b_\nu p_1^{5-\nu}$. (12)
B_n	$\dfrac{1}{n!}\cdot\dfrac{d^n B}{dp_1^n}$. (13)
c	Nir. (4)
c^*	Nur. (14)
C'	coma. (13)
γ_ν	terms of Γ_1 and Γ_2. (8)
Γ_ν	terms of the expansion of $\Delta\Lambda^*$. (5)
d	distance of diaphragm from preceding surface. (10)
d'	distance between successive surfaces. (15)

D	$\sum\limits_{\nu=1}^{f} \Delta\Lambda_\nu.$ (4)
D^*	$\sum\limits_{\nu=1}^{f} \Delta\Lambda_\nu.$ (5)
δ_0	$x-y+1.$ (7)
δ_ν	coefficient of $z^{2\nu+1}$ in the expansion of Δ. (7)
$\delta y_j, \delta u_j$	increments in y, u in the equations giving $\sin U_j$, etc. (5)
$\delta y_j', \delta u_j'$	increments in y, u in the equation giving $\sin U_j'$, etc. (5)
$\delta y_j^*, \ldots$	increments in $y, \ldots$ in the equations giving $\tan U_j, \ldots$ etc. (14)
Δ	$\Delta(\sin U + \sin I)$. (4)
ΔX_j	$X_j' - X_j.$ (1)
$\Delta^* X_j$	$X_{j+1} - X_j'.$ (1)
e	radius of aperture of diaphragm. (10)
f (suffix)	final surface. (5)
f'	posterior focal length. (10)
F'	paraxial focus (point conjugate to axial point of object). (3)
$g_n(x)$	coefficient of t^n in the expansion of $\sqrt{(1-t)(1-xt)}$. (6)
G_ν	terms of the expansion of Δ. (5)
h	height of object. (11)
h'	paraxial image height. (3)
H'	$(l_{of}' - L_f') \tan U_f'.$ (3)
i	paraxial angle of incidence. (1)
I	angle of incidence. (1)
θ	$c_q/c_p.$ (9)
k	$N/N'.$ (7)
l	paraxial axial intersection length. (1)
L	axial intersection length. (1)
λ	invariant of Lagrange type. (4)
Λ	$N(l-L)u \sin U.$ (4)
Λ^*	$N(l-L)u \tan U.$ (5)
N	refractive index. (1)
p (suffix)	quantities refering to the ray (1,0). (4)
p_ν	coefficients in the expansion of y of principal rays. (10)
q (suffix)	quantities referring to the ray (0,1). (4)
r	radius of surface. (1)
R, R'	arbitrary conjugate rays. (3)
R_0	associated ray of R. (4)
t_ν	coefficient of $u_{pf}'^{3-\nu}u^\nu$ in the expansion of $\tan U_f'$. (13)
t_n	current symbol in computing scheme. (15)
u	paraxial inclination of ray to axis. (1)
u	$\equiv u_1.$ (5)
U	inclination of ray to axis. (1)
v'	$H' - h'.$ (3)
w	semi-aperture of incident pencil of parallel rays. (13)
x	$(L-r)/r.$ (7)
y	$kx.$ (7)
y	$lu.$ (1)

y $\equiv y_1.$ (5)

z $\sin U.$ (7)

0 (suffix) quantities referring to focal paraxial ray. (5)

′ denotes any quantity behind surface corresponding to the "un-dashed" quantity before it. (1)

(α, β) a paraxial ray determined by $y_1 = \alpha$, $u_1 = \beta$. (4)

$[\alpha, \beta]$ a ray determined by $L_1 \sin U_1 = \alpha$, $\sin U_1 = \beta$. (4)

The quantities L, U, I, N, r, d' are the same as those used by Conrady (1929), so that the equations determining the passage of a ray through a surface are:

$$r \sin I = (L - r) \sin U, \qquad \ldots\ldots(1.1)$$

$$\Delta(N \sin I) = 0, \qquad \ldots\ldots(1.2)$$

$$\Delta(U + I) = 0, \qquad \ldots\ldots(1.3)$$

$$r \sin I' = (L' - r) \sin U', \qquad \ldots\ldots(1.4)$$

$$\Delta^* L = -d', \qquad \ldots\ldots(1.5)$$

and similar paraxial equations, i.e.

$$ri = (l - r)u, \qquad \ldots\ldots(1.6)$$

$$\Delta(Ni) = 0, \qquad \ldots\ldots(1.7)$$

$$\Delta(u + i) = 0, \qquad \ldots\ldots(1.8)$$

$$ri' = (l' - r)u', \qquad \ldots\ldots(1.9)$$

$$\Delta^* l = -d'. \qquad \ldots\ldots(1.10)$$

It should be noted that H' is positive if the point of intersection of R' with the plane of reference lies *below* the axis of the system. The phrase "correct to the nth order", as applied to any expression, is understood to mean that any errors involved in it are of at least the $(n+1)$th order.

§2. INTRODUCTION

It has been stated occasionally (Conrady, 1929, chap. 6 ; Hardy and Perrin, 1932, chap. 6) that the algebraic determination of the aberrations of lens systems other than the primary involves computations of such magnitude as to make such analysis useless from a practical point of view. Thus Conrady (p. 255) states: "... due consideration of the disturbing effects ... leads to the vast complication of the problem of determining the true secondary aberrations which up to the present has rendered the complete evaluation of the latter a task of almost hopeless complexity". Confining ourselves to the monochromatic tangential aberrations of a system of coaxial spherical refracting surfaces we shall show in the present paper (after suitably defining the "aberration of order n" of an arbitrary ray passing through the system) that the computation of the secondary aberrations represents relatively little labour in this case (see § 15, table 1); and that even the computation of the tertiary aberrations would not in practice be a task of insuperable magnitude. The methods used are essentially elementary in character; being entirely algebraic, the computations require no trigonometrical tracing, and are peculiarly adapted to the use of calculating machines. The simplicity of the method rests partly upon the choice of suitable variables, and partly upon a procedure which avoids the necessity for the explicit derivation of any summation theorems.

§3. DEFINITION OF THE ABERRATIONS

(a) Let an arbitrary ray R passing through the system from a given object point be defined by the values of $L \sin U$ and $\sin U$ at the first surface (see § 4 (b)). A paraxial ray from the axial point of the object (the latter being perpendicular to the axis) intersects the axis in the conjugate point F′ in the image space. The plane through F′ perpendicular to the axis (i.e. the paraxial focal plane) is intersected by the emerging ray R′, conjugate to R, at a perpendicular distance H' below the axis. If the system were perfect, all possible rays R′ from the given object point would have to pass through the ideal image point lying in the paraxial focal plane and a certain distance h' below F′, where h' is proportional to the height of the object. Let

$$v' = H' - h'. \qquad \ldots\ldots (3.1)$$

v' may simply be called the *aberration of the ray R*, v' being regarded as a function of the pair of variables $L_1 \sin U_1$ and $\sin U_1$. (Regarding the height of the object as variable, these two variables are independent.)   Now clearly, if the sign of U_1 be reversed, the signs of $\sin U_1$, $L_1 \sin U_1$, and of v' are all reversed.   Accordingly v' may be expanded in a series of homogeneous polynomials of the third, fifth,...., $(2n+1)$th degree $(n = 1, 2, 3, \ldots)$ in the independent variables $L_1 \sin U_1$, $\sin U_1$. We define the *aberration of order n* to be the polynomial of *degree* $(2n+1)$ in this expansion. Accordingly the primary aberrations involve four independent coefficients, the secondary six, and so on. Hence the determination of the aberrations up to and including the nth order requires the evaluation of $\sum_{\nu=1}^{n} (2\nu+2) = n(n+3)$ independent coefficients.

It may be useful slightly to extend the definition of v'. Instead of considering the intersection of R′ with the paraxial focal plane we may consider the intersection, H', with some arbitrary plane of reference (which is, of course, perpendicular to the axis). Then H' may also be expressed as a series of homogeneous polynomials of degree $(2n+1)$, $(n = 0, 1, 2, \ldots)$. We define v' in this case as H' minus the polynomial of the first degree. Our previous definition is contained in this as a special case.

(b) The four primary coefficients are, of course, intimately related to the primary aberrations of von Seidel. The actual relationships between them may be established by defining the spherical aberration, distortion, tangential coma and curvature of field purely geometrically and then expressing these in terms of v'. At the same time these definitions may, if desired, be retained for practical purposes on proceeding to higher orders. Accordingly we take as a measure of the

(i) *spherical aberration* the distance below the axis at which a ray from the axial point of the object strikes the paraxial focal plane;

(ii) *distortion* the value of v' corresponding to principal rays, i.e. rays passing through the centre of the diaphragm.

If the extreme rays of a pencil of rays be distinguished by the letters *a* and *b* respectively (the extreme rays being determined by the rim of the diaphragm), then

(iii) *coma* is measured by the length of the line drawn from the intersection of the *a* and *b* rays to the principal ray, parallel to the focal plane; whilst the

(iv) *curvature of field* is the perpendicular distance from the intersection of the a and b rays to the focal plane.

In the course of the evaluation of v' to the nth order, the direction of the ray in the image space is in any case determined to the $(n-1)$th order, so that all the data necessary for the calculations (i) to (iv) are available. By far the most important of these in practice is the coma, partly because of its connection with the question of unevenness of illumination in the image, and partly because it is not so easily found by inspection of the values of v' corresponding to different rays of a pencil (see § 5).

It should be observed that in the definition of v', as well as its computation, the position of the diaphragm is not involved. The variation of the aberrations with shift of diaphragm may therefore be investigated *after* the main computations have been carried out.

§4. ASSOCIATED RAYS

(*a*) With any given ray $R \equiv [L_1 \sin U_1, \sin U_1]$ we may associate a paraxial ray $R_0 \equiv (L_1 \sin U_1, \sin U_1)$, i.e. a paraxial ray the nominal initial data of which are the same as the actual data of R. We shall say that R and R_0 are *associated rays*. (The particular choice of the data determining R_0 is dictated by the equations of § 4 (*b*)).

Let α be any quantity connected with the paraxial ray R_0 which is linearly related to y_1 and u_1. Let A be a quantity connected with the associated ray R which reduces to α in the (paraxial) limit. Then clearly the difference $A - \alpha$ will be small, and it is easily seen that it must be of at least the third degree in y_1, u_1. (A is not thus uniquely determined; for instance, taking $\alpha \equiv u$, A might represent U, or $\sin U$, etc.)

Now one of the main difficulties of higher-order theory is the fact that we cannot simply replace any particular A by the corresponding α of the associated ray, as is done in primary theory, since such a procedure would make all but the primary terms meaningless. Accordingly it is desirable to find, if possible, expressions for the differences of the type $A - \alpha$, as explained above, for subsequent use, at the same time choosing the most convenient explicit form of A corresponding to the particular α.

In the following we define a *surface invariant* to be any quantity X_j obeying the equation

$$X_j' - X_j = \Delta X_j \equiv 0. \qquad (j = 1, 2 \ldots, f) \qquad \ldots\ldots(4.1)$$

X_j shall be called a *transfer invariant* if

$$X_{j+1} - X_j' = \Delta^* X_j \equiv 0; \qquad (j = 1, 2 \ldots, f-1) \qquad \ldots\ldots(4.2)$$

whilst if X_j satisfies both (4.1) and (4.2) simultaneously it shall be called a (true) *optical invariant*, for which evidently

$$X_i = X_j = X_k', \quad \text{for all } i, j, k. \qquad \ldots\ldots(4.3)$$

Moreover, it will be convenient to use the term *quasi-invariant* for any quantity which reduces to an optical invariant in the (paraxial) limit.

Now if we consider two *arbitrary* paraxial rays, say $(\bar{y}_1, \bar{u}_1)$ and (y_1, u_1), then the expression

$$\lambda = N(l - \bar{l})u\bar{u} \qquad \ldots\ldots(4.4)$$

is an optical invariant. For $\Delta * \lambda_j = 0$, obviously; and, at any surface,

$$\Delta\lambda = \Delta N\{(l-r)-(\bar{l}-r)\}u\bar{u} = \Delta Nr(\bar{i}u-iu), \text{ by } (1.6). \quad \ldots\ldots(4.5)$$

$\therefore$ since $\Delta Nri = \Delta c = 0$,

$$\Delta\lambda = c\Delta\left(\bar{u}-\bar{i}\frac{u}{i}\right) = c\Delta\left(\bar{u}+\bar{i}-\frac{i+u}{i}\bar{i}\right) = 0, \quad \ldots\ldots(4.6)$$

since $\Delta(u+i) = \Delta(\bar{i}/i) = 0$, by (1.7), (1.8).

Now let Λ be a quasi-invariant which reduces to λ in the limit. Then the fact that the expression

$$D = \sum_{\nu=1}^{f} \Delta\Lambda_\nu \quad \ldots\ldots(4.7)$$

does not vanish identically is implicitly connected with what may be called the imperfection of the optical system. A heuristic examination of such quasi-invariants may therefore be expected to yield a method whereby the desired differences, $A - \alpha$ (of which the difference $\Lambda - \lambda$ is itself merely a characteristic example), and thus the aberrations of the system, can be determined with reasonable ease.

(b) The equation (1.1) suggests the following explicit form of Λ:

$$\Lambda = N(l-L)u \sin U, \quad \ldots\ldots(4.8)$$

where (y_1, u_1) and $[L_1 \sin U_1, \sin U_1]$ are *arbitrary* rays. Then at any surface

$$\Delta\Lambda = c\Delta(\sin U + \sin I) = Ni\Delta(L \sin U), \quad \ldots\ldots(4.9)$$

since

$$\Delta\Lambda = \Delta N\{(l-r)-(L-r)\}u \sin U$$

$$= \Delta Nr(i \sin U - u \sin I), \text{ by } (1.1), (1.6)$$

$$= c\Delta\left(\sin U - \frac{u}{i}\sin I\right), \text{ by } (1.7)$$

$$= c\Delta\left(\sin U + \sin I - \frac{u+i}{i}\sin I\right),$$

which is (4.9), since $\Delta(u+i) = \Delta(\sin I/i) = 0$.

The great importance of equation (4.9) lies in its simplicity, in the sense that $\Delta\Lambda$ is split up into two factors, one of which depends upon y_1 and u_1 only, whilst the other factor is independent of these. Since the expression $\Delta(\sin I + \sin U)$ recurs frequently below, we shall simply write Δ for it, whenever confusion is not likely thereby to arise. Thus

$$\Delta\Lambda = c\Delta. \quad \ldots\ldots(4.10)$$

Equation (4.10) suggests that the adopted expression for Λ, (4.8), is the most convenient for our purpose.

Since, by definition, α depends linearly upon y_1 and u_1, we may write

$$\alpha_j = \alpha_{pj}y_1 + \alpha_{qj}u_1, \quad \ldots\ldots(4.11)$$

where the α_{pj}, α_{qj} are constants of the system. These constants are easily obtained in any given case. It is only necessary to trace the two (*paraxial*) rays (1,0) and (0,1) through the system. Then the actual value of α_j in the former is α_{pj}, whilst the value of α_j in the latter is α_{qj}.

We have, by (4.10), since $\Delta^*\Lambda_\nu = 0$,

$$N_j(l_j - L_j)u_j \sin U_j = N_1(l_1 - L_1)u_1 \sin U_1 + \sum_{\nu=1}^{j-1} c_\nu \Delta_\nu,$$

or

$$N_j y_j \sin U_j - N_j u_j L_j \sin U_j = N_1 y_1 \sin U_1 - N_1 u_1 L_1 \sin U_1 + \sum_{\nu=1}^{j-1} c_\nu \Delta_\nu.$$

$$\dots\dots(4.12)$$

Hence, since y_1 and u_1 are independent variables, (4.12) splits up, by virtue of (4.11), into the two equations

$$N_j y_{pj} \sin U_j - N_j u_{pj} L_j \sin U_j = N_1 \sin U_1 + \sum_{\nu=1}^{j-1} c_{pv} \Delta_\nu,$$

$$N_j y_{qj} \sin U_j - N_j u_{qj} L_j \sin U_j = -N_1 L_1 \sin U_1 + \sum_{\nu=1}^{j-1} c_{qv} \Delta_\nu.$$

$$\dots\dots(4.13)$$

These equations may be solved for $L_j \sin U_j$ and $\sin U_j$. Remembering that

$$\begin{vmatrix} y_{pj} & u_{pj} \\ y_{qj} & u_{qj} \end{vmatrix} = \frac{N_1}{N}, \qquad\qquad \dots\dots(4.14)$$

which follows at once by considering the optical invariant (4.4) for the special case of the two rays $(1,0)$ and $(0,1)$, we obtain the simple result

$$\sin U_j = u_{pj} L_1 \sin U_1 + u_{qj} \sin U_1 + \sum_{\nu=1}^{j-1} \frac{1}{N_1}(c_{pv} u_{qj} - c_{qv} u_{pj})\Delta_\nu,$$

$$\dots\dots(4.15)$$

and analogous equations for $L_j \sin U_j$, $\sin I_j$, $\sin I_j'$, the only change being that u_{pj} is replaced by y_{pj}, i_{pj}, i_{pj}' respectively, and similarly for u_{qj}. Equation (4.15) may be rewritten more simply in terms of the associated ray of $[L_1 \sin U_1, \sin U_1]$, i.e.

$$\sin U_j = u_j + \sum_{\nu=1}^{j-1} \frac{1}{N_1}(c_{pv} u_{qj} - c_{qv} u_{pj})\Delta_\nu, \qquad \dots\dots(4.16)$$

and analogous equations for $L_j \sin U_j$, $\sin I_j$, $\sin I_j'$. Notice that the equations for $\sin U_j'$ and $L_j' \sin U_j'$ are similar, except that the summations are extended from $\nu = 1$ to $\nu = j$ instead of to $j - 1$.

We have thus found the desired equations giving the differences between corresponding quantities of associated rays.

§5. PRINCIPLES OF THE PRESENT METHOD

(a) Writing $\Lambda^* = \Lambda \sec U$ we have, since $\Delta^*\Lambda^* = 0$,

$$\Lambda_j^* = \Lambda_1^* + \sum_{\nu=1}^{f} \Delta\Lambda_\nu^*. \qquad\qquad \dots\dots(5.1)$$

Adopting the restricted definition of v' from §3 it follows, on taking the paraxial ray contained in Λ^* to pass through the axial point of the object, that

$$\Lambda_j^{*\prime} = N' u_{0f}'(l_{0f}' - L') \tan U_j' = N_j' u_{0f}' H' \qquad \dots\dots(5.2)$$

and

$$\Lambda_1^* = N_1 u_{01} H_1 = N_1 u_{01} h = N_j' u_{0f}' h', \qquad \dots\dots(5.3)$$

where h is the height of the object, so that

$$N_j' u_{0f}' v' = \sum_{\nu=1}^{f} \Delta\Lambda_\nu^* = D^*, \text{ say.} \qquad \dots\dots(5.4)$$

(We have distinguished quantities referring to this particular paraxial ray by the suffix 0, and we shall refer to it in future as the " focal paraxial ray ".) Accordingly we may take the "contribution by the jth surface to the final aberration" to be $\Delta\Lambda_j^*/N_j'u_0'_f$.

Now suppose (dropping the suffixes referring to the particular surface) $\Delta\Lambda^*$ to be expanded in ascending powers of $\sin U$ (see § 8), say

$$\Delta\Lambda^* = \Gamma_1 + \Gamma_2 + \Gamma_3 + \dots, \qquad (5.5)$$

where Γ_n is the term in $\sin^{2n+1} U$. Γ_n may be rewritten by absorbing $\sin^{2n+1} U$ in the factor that multiplies it. We have then Γ_n in the form of the sum of a number of terms each of which consists of the product of $2n+1$ sines of one or more of the angles U, I, I', as well as of certain quantities referring to the focal paraxial ray. Replacing the sines formally by the corresponding quantities of the associated ray, the latter being regarded as functions of y_1 and u_1, we are led to the series

$$\Gamma_1(y,u) + \Gamma_2(y,u) + \Gamma_3(y,u) + \dots, \qquad \dots (5.6)$$

where $\Gamma_n(y,u)$ is a homogeneous polynomial of degree $2n+1$ in y, u. (We have simply written y, u for y_1, u_1, respectively, and shall do so below whenever this procedure is not likely to cause confusion.)

Now clearly the errors committed in this process cause $\Gamma_2, \Gamma_3, \dots$ to lose their definite meaning for our purpose since the terms neglected in Γ_1 are of the fifth and higher degree, i.e. of the same (and higher) degree as Γ_2, and so on. However, returning to equation (4.16), we see that it may be written

$$\sin U_j = u_{pj}\left(y - \frac{1}{N_1}\sum_{\nu=1}^{j-1} c_{q\nu}\Delta_\nu\right) + u_{qj}\left(u + \frac{1}{N_1}\sum_{\nu=1}^{j-1} c_{p\nu}\Delta_\nu\right). \qquad \dots (5.7)$$

Hence, putting

$$\left.\begin{aligned}\delta y_j &= -\frac{1}{N_1}\sum_{\nu=1}^{j-1} c_{q\nu}\Delta_\nu, \\ \delta u_j &= +\frac{1}{N_1}\sum_{\nu=1}^{j-1} c_{p\nu}\Delta_\nu\end{aligned}\right\} \qquad \dots (5.8)$$

and $\delta y_j'$, $\delta u_j'$ similarly with the summations extended from $\nu=1$ to j instead of $j-1$, we have

$$A_j = \alpha_{pj}(y + \delta y_j) + \alpha_{qj}(u + \delta u_j), \qquad \dots (5.9)$$

where A stands for *any one* of the quantities $L\sin U$, $\sin U$, $\sin I$, $\sin I'$, and each α corresponds to the particular A. When A stands for $L'\sin U'$ or $\sin U'$, δy and δu must be replaced by $\delta y'$ and $\delta u'$ respectively in (5.9).

The important feature of (5.9) is that the value of any of the quantities which A represents is given by an equation closely analogous to the paraxial linear equation (4.11), i.e. by giving y and u on the right-hand side of the latter the *same* increments δy and δu, *irrespective* of which of the four alternative expressions A happens to represent. Hence, assuming δy_j and δu_j to be expressed as functions of y and u correct to the $(n-1)$th order (see below), we need only substitute $y + \delta y_j$ and $u + \delta u_j$ for y and u respectively in (5.6), rejecting all terms of degree greater than $2n+1$, in order to make that series represent $\Delta\Lambda_j^*$ exactly to the nth order. The terms of the resulting series may then be grouped together so as to give rise to a series of homogeneous polynomials in y, u of degree

3, 5, ... $2n+1$. Then the polynomial of degree $2m+1$ is the contribution (except for the constant factor $1/N'u'_j$) by the jth surface to the aberration of order m.

(b) We have yet to show how the δy_j and δu_j are to be determined as functions of y and u. The simplest way of doing this, both in theory and in practice, is by means of a step-by-step method, following closely the procedure outlined above. For this purpose we imagine Δ to be expanded (at any surface) in a series exactly analogous to (5.5) (see §7), i.e.

$$\Delta = G_1 + G_2 + G_3 + \dots, \qquad \dots\dots(5.10)$$

which again leads to a derived series, similar to (5.6), i.e.

$$G_1(y, u) + G_2(y, u) + G_3(y, u) + \dots. \qquad \dots\dots(5.11)$$

Rejecting all but G_1, δy_j and δu_j may be calculated correct to the first order (i.e. 3rd degree) by means of (5.8); thus

$$\left.\begin{aligned}
\delta y_j &= -\frac{1}{N_1} \sum_{\nu=1}^{j-1} c_{q\nu} G_{1\nu}(y, u), \\
\delta u_j &= +\frac{1}{N_1} \sum_{\nu=1}^{j-1} c_{p\nu} G_{1\nu}(y, u).
\end{aligned}\right\} \qquad \dots\dots(5.12)$$

Hence from (5.11) we can write down, using the expressions (5.12) at every surface, the terms

$$G_{1j}(y + \delta y_j, u + \delta u_j) + G_{2j}(y, u), \qquad \dots\dots(5.13)$$

rejecting all terms of degree higher than the fifth. (5.13) is then the expansion of Δ_j correct to the second order. This allows us immediately to calculate δy_j and δu_j correct to the second order by substituting in (5.8). Hence with these values of δy_j and δu_j,

$$G_{1j}(y + \delta y_j, u + \delta u_j) + G_{2j}(y + \delta y_j, u + \delta u_j) + G_{3j}(y, u) \quad \dots\dots(5.14)$$

(rejecting all terms of degree greater than the seventh) represents Δ_j correct to the third order. This process may be obviously repeated until we reach the expansion of Δ_j correct to the $(n-1)$th order. (Notice that, of course, the only new terms to be calculated on passing from order m to order $m+1$ are those of order $m+1$, all terms of lower order remaining unchanged.) Then, finally,

$$\sum_{\nu=1}^{n} \Gamma_{\nu j}(y + \delta y_j, u + \delta u_j) \qquad \dots\dots(5.15)$$

(rejecting all terms of degree higher than the $(2n+1)$th) is the desired expansion of $\Delta\Lambda_j^*$ correct to the nth order.

Alternatively we might calculate $\delta y'_j$ and $\delta u'_j$ correct to the nth order and then directly determine $\Lambda_j^{*'}$. Though this method might be somewhat shorter in practice (particularly for large n), it would not yield directly the separate contributions by each surface to the final aberration.

§6. g-POLYNOMIALS

The explicit expansion of Δ is greatly facilitated by the use of certain simple polynomials $g_n(x)$ defined by

$$\sqrt{(1-t)(1-xt)} = \sum_{n=0}^{\infty} g_n(x) t^n, \qquad \dots\dots(6.1)$$

where x, t are independent variables whose range is suitably restricted so as to ensure the convergence of (6.1).

Replacing t by t/x in (6.1)

$$\sqrt{(1-t/x)(1-t)} = \sum_{n=0}^{\infty} g_n(x)x^{-n}t^n \equiv \sum_{n=0}^{\infty} g_n(1/x)t^n.$$

$$\therefore \quad g_n(x) = x^n g_n(1/x).$$

Hence $\qquad\qquad y^n g_n(x/y) \equiv x^n g_n(y/x) \qquad\qquad \ldots\ldots(6.2)$

is a polynomial symmetrical in x and y.

Expanding the square roots on the left-hand side of (6.1), it follows easily that

(or, also,

$$g_n(x) = (-1)^n \sum_{\nu=0}^{n} \binom{\frac{1}{2}}{n-\nu}\binom{\frac{1}{2}}{\nu}x^\nu$$

$$\left.\begin{array}{c}\\ \\ \end{array}\right\} \ldots\ldots(6.3)$$

$$g_n(x) = (-1)^n \binom{\frac{1}{2}}{n}F(-n, -\tfrac{1}{2}; \tfrac{3}{2}-n; x),$$

where $F(\alpha, \beta; \gamma; x)$ denotes the hypergeometric function of Gauss).

Setting $x = +1$ in (6.1) we have at once

$$g_n(+1) = 0, \qquad\qquad (n>1).$$

Similarly $\qquad\quad g_n(-1) = 0, \qquad\qquad (n \text{ odd}).$ $\qquad \left.\begin{array}{c}\\ \end{array}\right\} \ldots\ldots(6.4)$

The recurrence formula satisfied by the $g_n(x)$ is

$$(n+1)g_{n+1}(x) = (n-\tfrac{1}{2})(x+1)g_n(x) - (n-2)x g_{n-1}(x). \qquad \ldots\ldots(6.5)$$

In particular

$$\begin{array}{ll}
g_0(x) = 1, & g_1(x) = -\tfrac{1}{2}(x+1), \\
g_2(x) = -\tfrac{1}{8}(x-1)^2, & g_3(x) = -\tfrac{1}{16}(x-1)^2(x+1), \\
g_4(x) = -\tfrac{1}{128}(x-1)^2(5x^2+6x+5), \\
g_5(x) = -\tfrac{1}{256}(x-1)^2(x+1)(7x^2+2x+7).
\end{array} \quad \left.\begin{array}{c}\\ \\ \\ \\ \end{array}\right\} \ldots\ldots(6.6)$$

§7. EXPANSION OF Δ

We have

$$\Delta = \sin U' + \sin I' - \sin U - \sin I,$$

which, by (1.3),

$$= \cos U \sin I \cos I' - \cos U \cos I \sin I' + \sin U \cos I \cos I'$$
$$+ \sin U \sin I \sin I' - \sin I + \sin I' - \sin U.$$

$$\ldots\ldots(7.1)$$

For convenience, in this paragraph, let

$$(L-r)/r = x; \quad kx = Nx \quad N' = y; \quad \sin U = z,$$

so that

$$\sin I = xz; \quad \sin I' = yz. \qquad \left.\begin{array}{c}\\ \end{array}\right\} \ldots\ldots(7.2)$$

Therefore, since $\cos I = (1-\sin^2 I)^{\frac{1}{2}}$, etc.,

$$\Delta = z\{x\sqrt{(1-z^2)(1-y^2z^2)} - y\sqrt{(1-z^2)(1-x^2z^2)} $$
$$+ \sqrt{(1-x^2z^2)(1-y^2z^2)} + xyz^2 - x + y - 1\}, \qquad \ldots\ldots(7.3)$$

or, in terms of the polynomials of §7,

$$\Delta = z\{xyz^2 - x + y - 1 + \sum_{n=0}^{\infty} [xg_n(y^2) - yg_n(x^2) + x^{2n}g_n(y^2/x^2)]z^{2n}\}. \quad \ldots\ldots(7.4)$$

$$\therefore \quad \Delta = \sum_{n=1}^{\infty}{}^{*} \delta_n z^{2n+1}, \quad\quad\quad \ldots\ldots(7.5)$$

where

$$\delta_1 = \tfrac{1}{2}(x-y)(x+1)(y-1),$$
$$\delta_2 = \tfrac{1}{8}(x-y)(x+1)(y-1)(1-x+y+x^2+xy+y^2),$$
$$\delta_3 = \tfrac{1}{16}(x-y)(x+1)(y-1)(1-x+y+x^2+y^2-x^3+y^3+x^4+x^3y \\ +x^2y^2+xy^3+y^4),$$

$$\left. \vphantom{\begin{matrix}1\\1\\1\\1\end{matrix}} \right\} \ldots\ldots(7.6)$$

and generally

$$\delta_n = xg_n(y^2) - yg_n(x^2) + x^{2n}g_{\hat{n}}(k^2), \quad n \geqslant 2.$$

Writing $\delta_0 = x - y + 1$, we have in particular

$$\delta_2 = \tfrac{1}{8}\delta_1\{3(x^2+y^2+1) - \delta_0{}^2\}. \quad\quad \ldots\ldots(7.7)$$

It is evident that every $\delta_n(n>1)$ has δ_1 as a factor, since δ_n vanishes when $x = -1$, $y = +1$, and $k = 1$, in virtue of (6.4).

Absorbing z^{2n+1} in δ_n we obtain the explicit form of the $G_1, G_2, \ldots$ of (5.10); thus

$$G_1 = \tfrac{1}{2}(\sin I + \sin U)(\sin I - \sin I')(\sin I' - \sin U),$$
$$G_2 = \tfrac{1}{4}G_1(\sin^2 U - \sin U \sin I + \sin U \sin I' + \sin^2 I \\ + \sin I \sin I' + \sin^2 I'),$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad ., \text{etc.}$$

$$\left. \vphantom{\begin{matrix}1\\1\\1\end{matrix}} \right\} \ldots\ldots(7.8)$$

To obtain the $G_\nu(y, u)$ of (5.11) we need only replace $\sin U, \sin I, \sin I'$ by the values of u, i, i' of the associated ray, giving

$$G_1(y, u) = \tfrac{1}{2}(i+u)(i-i')(i'-u),$$
$$G_2(y, u) = \tfrac{1}{8}G_1(y, u)\{3(i^2+u^2+i'^2) - u'^2\}, \text{ by } (7.7),$$
$$G_3(y, u) = \tfrac{1}{8}G_1(y, u)(u^4 - iu^3 + i'u^3 + i^2u^2 + i'^2u^2 - i^3u + i'^3u + i^4 \\ + i^3i' + i^2i'^2 + ii'^3 + i'^3)$$
$$= \tfrac{1}{2}u^2G_2(y, u) + \tfrac{1}{8}G_1(y, u)\left\{u(i'^3 - i^3) - ii'u^2 + \frac{i^5 - i'^5}{i - i'}\right\}$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad ., \text{etc.}$$

$$\left. \vphantom{\begin{matrix}1\\1\\1\\1\\1\end{matrix}} \right\} \ldots\ldots(7.9)$$

where $i, u, \ldots$ are given by (4.11). Thus, for instance,

$$G_1(y, u) = \tfrac{1}{2}\{(i+u)_p v_1 + (i+u)_q u_1\}\{(i-i')_p y_1 \\ + (i-i')_q u_1\}\{(i'-u)_p y_1 + (i'-u)_q u_1\}. \quad \ldots\ldots(7.10)$$

The case of a plane surface is easily disposed of. It is only necessary to calculate rG_ν in place of G_ν (so that instead of the vanishing factor $(i+u)$ in G_ν we now have $r(i+u) = y$, which remains finite), and to allow for this by substituting Ni_p for c_p and Ni_q for c_q at this surface in (5.8), etc.

§8. EXPANSION OF $\Delta\Lambda^*$

We have

$$\Delta\Lambda^* = \Delta(\Lambda \sec U) = \sec U\Delta\Lambda + \Lambda\Delta \sec U + (\Delta\Lambda)(\Delta \sec U),$$
$$\therefore \quad \Delta\Lambda^* = \sec U'\Delta\Lambda + \Lambda\Delta \sec U. \quad\quad \ldots\ldots(8.1)$$

If we desire to carry out the computations to a high order, much labour will evidently be saved if we express the Γ_ν as far as possible in terms of the G_ν since the latter are to be calculated in any case. This may be done as follows:

By (7.5), remembering that $\delta_0 = x - y + 1$,

$$\sin U' = \sum_{\nu=0}^{\infty} \delta_\nu z^{2\nu+1}; \qquad \ldots\ldots(8.2)$$

also

$$\sec U = (1 - \sin^2 U)^{-\frac{1}{2}}$$

and

$$\left.\begin{aligned}
\therefore \quad \sec U &= 1 + \tfrac{1}{2}z^2 + \tfrac{3}{8}z^4 + \tfrac{5}{16}z^6 + \ldots \\
\sec U' &= 1 + \tfrac{1}{2}\delta_0{}^2 z^2 + (\delta_0\delta_1 + \tfrac{3}{8}\delta_0{}^4)z^4 \\
&\quad + (\tfrac{1}{2}\delta_1{}^2 + \delta_0\delta_2 + \tfrac{3}{2}\delta_0{}^3\delta_1 + \tfrac{5}{16}\delta_0{}^6)z^6 + \ldots
\end{aligned}\right\} \qquad \ldots\ldots(8.3)$$

$$\begin{aligned}
\therefore \Delta\Lambda^* &= c_0\{1 + \tfrac{1}{2}\delta_0{}^2 z^2 + (\delta_0\delta_1 + \tfrac{3}{8}\delta_0{}^4)z^4 + \ldots\}\{\delta_1 z^3 + \delta_2 z^5 + \delta_3 z^7 + \ldots\} + \Lambda\Delta \sec U \\
&= c_0\{\delta_1 z^3 + (\delta_2 + \tfrac{1}{2}\delta_0{}^2\delta_1)z^5 + (\delta_3 + \tfrac{1}{2}\delta_0{}^2\delta_2 + \delta_0\delta_1{}^2 + \tfrac{3}{8}\delta_0{}^4\delta_1)z^7 + \ldots\} + \Lambda\Delta \sec U.
\end{aligned}$$

$$\ldots\ldots(8.4)$$

Combining the various powers of z with the factors which multiply them, and replacing the resulting sines of the angles by the corresponding quantities of the associated ray, we obtain

$$\left.\begin{aligned}
\Gamma_1 &= c_0 G_1 + \tfrac{1}{2}\lambda_0(u'^2 - u^2) \\
\Gamma_2 &= c_0(G_2 + \tfrac{1}{2}u'^2 G_1) + \lambda_0(u'G_1 + \tfrac{3}{8}\overline{u'^4 - u^4}) \\
\Gamma_3 &= c_0(G_3 + \tfrac{1}{2}u'^2 G_2 + u'G_1{}^2 + \tfrac{3}{8}u'^4 G_1) + \lambda_0(u'G_2 + \tfrac{1}{2}G_1{}^2 \\
&\qquad\qquad\qquad\qquad\qquad\qquad + \tfrac{3}{2}u'^3 G_1 + \tfrac{5}{16}\overline{u'^6 - u^6})
\end{aligned}\right\} \qquad \ldots\ldots(8.5)$$

$$\cdots\cdots\cdots\cdots\cdots, \text{ etc.}$$

In practice Γ_1 and Γ_2 are particularly important. Let us consider these in particular. We have by (8.5), (7.9)

$$\Gamma_1 = \tfrac{1}{2}c_0(i+u)(i-i')(i'-u) + \tfrac{1}{2}(i-i')(u+u')\lambda_0.$$

Also

$$\begin{aligned}
G_2 + \tfrac{1}{2}u'^2 G_1 &= G_1\{\tfrac{1}{2}u'^2 + \tfrac{3}{8}(u^2 + i^2 + i'^2) - \tfrac{1}{8}u'^2\} \\
&= \tfrac{3}{8}G_1(u^2 + i^2 + i'^2 + u'^2)
\end{aligned}$$

and

$$G_1 + \tfrac{3}{8}(u'^4 - u^4) = \tfrac{1}{8}(i - i')\{4u'(i+u)(i'-u) + 3(u+u')(u^2 + u'^2)\},$$

which may be rearranged, giving

$$u'G_1 + \tfrac{3}{8}(u'^4 - u^4) = \tfrac{1}{8}(i - i')\{\tfrac{3}{2}(u+u')(u^2 + i^2 + i'^2 + u'^2) + (i+u)(i-i')(u' - 3u)\}.$$

Hence if we write

$$\left.\begin{aligned}
\gamma_1 &= \tfrac{1}{2}c_0(i+u)(i-i')(i'-u), \\
\gamma_2 &= \tfrac{1}{2}(i-i')(u+u')\lambda_0 = \tfrac{1}{2}(u'^2 - u^2)\lambda_0, \\
\gamma_3 &= \tfrac{3}{8}(u^2 + i^2 + i'^2 + u'^2), \\
\gamma_4 &= \tfrac{1}{4}(u' - 3u)\lambda_0/c_0,
\end{aligned}\right\} \qquad \ldots\ldots(8.6)$$

we have finally

$$\left.\begin{aligned}
\Gamma_1 &= \gamma_1 + \gamma_2, \\
\Gamma_2 &= (\gamma_1 + \gamma_2)\gamma_3 + \gamma_1\gamma_4,
\end{aligned}\right\} \qquad \ldots\ldots(8.7)$$

which is the form most convenient for computation. (Notice that at every surface λ_0 may be replaced by λ_{01} at once.)

Correct to the second order, the contribution by the jth surface to $N_f' u_{0f}' v'$ is then given by

$$(\gamma_{1j} + \gamma_{2j})$$

$$+ \left\{ (\gamma_{1j} + \gamma_{2j})\gamma_{3j} + \gamma_{1j}\gamma_{4j} + \frac{1}{N_1} \frac{\partial(\gamma_{1j} + \gamma_{2j})}{\partial u_1} \sum_{\nu=1}^{j-1} \frac{c_{p\nu}}{c_{0\nu}} \gamma_{1\nu} - \frac{1}{N_1} \frac{\partial(\gamma_{1j} + \gamma_{2j})}{\partial y_1} \sum_{\nu=1}^{j-1} \frac{c_{q\nu}}{c_{0\nu}} \gamma_{1\nu} \right\}$$

$$\ldots \ldots (8.8)$$

§9. CHANGE OF POSITION OF OBJECT OR OF FOCAL PLANE

Since the ray RR' (§3) is quite arbitrary, it follows that the position of the object enters into the Γ_ν of §8 only through the coefficients c_0 and λ_0. In fact, using the extended definition of v' of §3, we are in every case only interested in the perpendicular distance from the principal axis of the point in which R' intersects a chosen focal plane, the position of which is defined by l_{0f}', where the suffix 0 is now used to refer to a paraxial ray having the required l_{0f}' as its final axial intersection length. That is, if we are interested in the performance of the optical system for widely differing positions of the object (or, more generally, for widely differing positions of the plane of reference), we need only compute v_p' and v_q' separately (i.e. v' corresponding to the rays $(y_{01}, u_{01}) \equiv (1,0)$ and $(0,1)$ respectively). For then, by (5.4),

$$N_f' u_{0f}' v' = y_{01} \sum_{j=1}^{f} \Delta \Lambda_{pj}^* + u_{01} \sum_{j=1}^{f} \Delta \Lambda_{qj}^*. \qquad \ldots \ldots (9.1)$$

Here y_{01} and u_{01} are connected by the condition that $y_{0f}/u_{0f}' = l_{0f}'$. Hence writing (9.1) in the form

$$u_{0f}' v' = y_{01} u_{pf}' v_p' + u_{01} u_{qf}' v_q', \qquad \ldots \ldots (9.2)$$

and applying (4.11) and (4.14), it follows that

$$v' = \frac{N_f'}{N_1} \{ (l_{0f}' u_{qf}' - y_{qf})u_{pf}' v_p' - (l_{0f}' u_{pf}' - y_{pf})u_q' v_q' \}. \qquad \ldots \ldots (9.3)$$

According to this equation, if the reference plane be moved through a distance $\delta l_{0f}'$, v' changes by the amount

$$\delta v' = \frac{N_f'}{N_1} u_{pf}' u_{qf}'(v_p' - v_q')\delta l_{0f}'. \qquad \ldots \ldots (9.4)$$

Notice that the actual magnitude of $\delta l_{0f}'$ is not restricted to be small.

The amount of labour involved in calculating both v_p' and v_q' is not much greater than that involved in calculating v' alone. We shall indicate how this is to be done most easily, considering only primary and secondary terms. (Higher orders are not essentially different.) Thus, from (8.6) we first determine in the ordinary way the quantities γ_1, γ_2, γ_3 and γ_4 with $(y_{01}, u_{01}) \equiv (1,0)$, i.e. we determine the four quantities γ_{1p}, γ_{2p}, γ_{3p}, γ_{4p} and hence the three products $\gamma_{1p}\gamma_3$, $\gamma_{2p}\gamma_3$ and $\gamma_{1p}\gamma_{4p}$.

Let $c_q/c_p = \theta$. Then, since

$$\lambda_q = -\gamma_1 \lambda_p/u_1, \qquad \ldots \ldots (9.5)$$

we have

$$
\left.
\begin{aligned}
\gamma_{1q} &= \theta\gamma_{1p}, \\
\gamma_{2q} &= -y_1\gamma_{2p}/u_1, \\
\gamma_{1q}\gamma_3 &= \theta\gamma_{1p}\gamma_3, \\
\gamma_{2q}\gamma_3 &= -y_1\gamma_{2p}\gamma_3/u_1, \\
\gamma_{1q}\gamma_{4q} &= -y_1\gamma_{1p}\gamma_{4p}/u_1.
\end{aligned}
\right\} \quad \dots\dots(9.6)
$$

(Notice that the δy_j already involve γ_{1q}.) Thus, having determined v_p', it will be seen (remembering that the δy_j and δu are unchanged) that the only new multiplications required are those of the coefficients of $\gamma_{1p}\gamma_3$ by θ; apart from these only a number of additions is required.

§10. THE DIAPHRAGM

A pencil of rays is frequently specified by the position of the object point and the position and aperture of the diaphragm. The "extreme rays" of the pencil (assuming that there is no vignetting) are the rays grazing the rim of the diaphragm. (The aperture of the latter is supposed to be circular.) For a given position of the diaphragm there exists therefore a definite relationship between the values of y and u (i.e. y_1 and u_1) specifying principal rays. To determine this let the diaphragm be situated a distance d behind the surface immediately preceding it, say the jth. Then, paraxially, the desired relationship is given by

$$
l \equiv d = \frac{y_{pj}y + y_{qj}'u}{u_{pj}'y + u_{qj}'u}, \qquad \dots\dots(10.1)
$$

$$
y = \frac{u_{qj}'d - y_{qj}}{y_{pj} - u_{pj}'d} = p_1 u, \text{ say.} \qquad \dots\dots(10.2)
$$

(p_1 is simply the paraxial location of the centre of the entrance pupil with respect to the first surface.) However, we have strictly

$$
L' \equiv d = \frac{y_j + y_{pj}\delta y_j' + y_{qj}\delta u'}{u' + u_{pj}'\delta u' + u_{qj}'\delta u'},
$$

which may be written

$$
y = p_1 u - \delta y_j' + p_1 \delta u_j'. \qquad \dots\dots(10.3)
$$

To the first order the $\delta y_j'$ and $\delta u_j'$ may be calculated by setting $y = p_1 u$, ($\delta y_j'$ and $\delta u_j'$ have already been computed for *general* values of y_1 and u_1), the error so committed being only of the second order. Hence, correct to the second order,

$$
y = p_1 u + p_3 u^3, \qquad \dots\dots(10.4)
$$

where p_3 is a polynomial of the forth degree in p_1. This process may be continued to any desired order. Writing (10.4) in the form

$$
y/u - p_1 = (L_{pr})_1 - p_1' = p_3 u^2 + \dots, \qquad \dots\dots(10.5)
$$

we see that the right-hand side of (10.5) gives the variation of the location with respect to the first surface of the centre of the entrance pupil, due to the imperfections of the optical system, as a function of the inclination to the axis of the incident principal ray.

Similarly the relation between y and u for the extreme rays, corresponding to (10.2), is

$$y = p_1 u \pm \frac{e}{y_p - u_p' d},\qquad\ldots\ldots(10.6)$$

where e is the radius of the aperture of the diaphragm. Higher approximations, corresponding to (10.5), may of course be determined.

In the present notation, the stop-number of the system, defined as the ratio of the posterior focal length $f'(=1/u_{pf}')$ to the diameter of the entrance pupil, is given by

$$\text{stop number} = \frac{y_{pj} - u_{pj}' d}{2 u_{pf}' e}.\qquad\ldots\ldots(10.7)$$

§ 11. v', AND SPHERICAL ABERRATION

(a) Let us assume that v' has been computed (§§5–8) and is available in the final form

$$v' = a_0 y^3 + a_1 y^2 u + a_2 yu^2 + a_3 u^3 + b_0 y^5 + b_1 y^4 u + b_2 y^3 u^2$$
$$+ b_3 y^2 u^3 + b_4 yu^4 + b_5 u^5 + \text{higher terms}.\qquad\ldots\ldots(11.1)$$

(11.1) here refers to the last surface of the optical system considered (which may of course be any intermediate surface of an actual system), the coefficients being evaluated for a definite position of the object. We shall take the plane of reference to be the paraxial focal plane.

To gain information about the system, which can be easily visualized, it will frequently be desirable (though by no means essential) to plot v' as a function of y and u. If the object is situated at infinity it is advisable to plot v' as a function of y for various fixed values of u, whilst if the object, of height h, is situated at l_{01}, one may plot v' against u for various fixed values of h; y and u in this case are related by the condition

$$(l_{01} - L_1) \tan U_1 = \text{const.} = h,$$

or in terms of the associated ray,

$$y = l_{01} u - h\sqrt{(1 - u^2)}.\qquad\ldots\ldots(11.2)$$

The position of the diaphragm may often be decided upon by inspection of the graphs so obtained.

(b) If the object is situated at infinity (as is frequently the case with photographic objectives) the *spherical aberration* is given by (11.1) on making $u = 0$, i.e.

$$\text{spher. ab.} = (v')_{u=0} = a_0 y^3 + b_0 y^5 + \ldots.\qquad\ldots\ldots(11.3)$$

This equation gives the radius of the disc of confusion projected by the system on a screen placed in the position of the focal plane. In practice we may also measure the spherical aberration by the value of the expression

$$[(l_{0f}' - L_f') \sin U_f']_{u=0},\qquad\ldots\ldots(11.4)$$

which is more convenient from the point of view of the computation of higher orders.

Thus for the *tertiary spherical aberration* we require essentially the calculation of $G_1(y, u)$ omitting the term in u^3; of $G_2(y, u)$ rejecting all terms except those in

y^5 and y^4u; and of the term in y^7 in $G_3(y, u)$. (Since the $\delta y_j'$ and $\delta u_j'$ are then in any case known to the second order, $\sin U_j'$, and hence $(v')_{u=0}$ may, if desired, be calculated subsequently directly from (11.4). The difference between these results is, however, in general, quite negligible.)

The great simplicity apparent in the calculation of the higher-order spherical aberration *ab initio*, as indicated above, is there due to the fact that one of the independent variables, namely u, is zero. Hence to obtain the same simplicity when the object is not at infinity but at l_{01}, say, we must proceed somewhat differently, viz. by introducing the independent variable

$$\bar{y}_1 = y_1 - l_{01}u_1 \qquad \dots\dots(11.5)$$

in place of y_1. Equation (5.7) then becomes

$$\sin U_j = u_{pj}\left(\bar{y} - \sum_{v=1}^{j-1} \frac{1}{N_1}(c_{qv} + l_{01}c_{pv})\Delta_v\right) + (u_{qj} + u_{pj}l_{01})\left(u + \sum_{v=1}^{j-1}\frac{1}{N_1}c_{pv}\Delta_v\right).$$
$$\dots\dots(11.6)$$

We have therefore again a set of equations of the type (5.8), (5.9), etc., except that where previously any quantity referring to the ray $(0,1)$ appeared we now have in its place the corresponding quantity referring to the ray $(l_{01}, 1)$. Accordingly the computations now begin with the tracing of the two rays $(l_{01},1)$ and $(1,0)$. And since one of the independent variables is now zero, the calculations are similar in every detail to those in the case where the object point is situated at infinity.

If, as will often be the case, the primary and secondary terms are already known, it may nevertheless be worth while to calculate them again by the method above. For in this way a good check is obtained on equation (11.1), from which the spherical aberration is obtained by setting $y = l_{01}u$, so that *all* the coefficients appearing in it are checked in this way.

(c) In practice one will generally limit oneself to the computation of the exact primary and secondary aberrations; partly because the information so obtained will be sufficient in many cases, and partly because the evaluation of the *general* tertiary (and higher) terms is rather lengthy. As we have seen, the computation of the tertiary spherical aberration is, however, relatively simple. Consequently it will frequently be advisable to carry it out. For obviously (taking the object to be at infinity) when the inclination of the incident pencil (i.e. u_1) is sufficiently small, the difference between the actual aberration and the sum of the primary and secondary terms is in general dominated by the next higher term in (11.1) not involving u; the latter being the tertiary spherical aberration. Similarly when l_{01} is finite.

If the plane of reference is other than the paraxial focal plane, the considerations above are not essentially altered. Thus, if the system were perfect, the radius of the disc of confusion would be given by a simple term involving y and u linearly. In the actual system the radius of the disc of confusion is, therefore, the sum of this linear term and v' (extended definition).

§12. DISTORTION

From the definition of §3 it will be seen that the distortion is represented by v' when y is replaced by its value in terms of u as given by (10.4). We shall

restrict ourselves in this and the next paragraph to the first and second orders, the extension to all orders being obvious. Thus .

$$\text{Distortion} = (a_0 p_1{}^3 + a_1 p_1{}^2 + a_2 p_1 + a_3)u^3$$
$$+ \{(b_0 p_1{}^5 + b_1 p_1{}^4 + b_2 p_1{}^3 + b_3 p_1{}^2 + b_4 p_1 + b_5) + p_3(3a_0 p_1{}^2 + 2a_1 p_1 + a_2)\}u^5.$$
$$\dots\dots(12.1)$$

Therefore, if

$$\left. \begin{aligned} A &= \overset{3}{\underset{\nu=1}{\Sigma}} a_\nu p^{3-\nu} \\[4pt] B &= \overset{5}{\underset{\nu=1}{\Sigma}} b_\nu p^{5-\nu}, \end{aligned} \right\} \dots\dots(12.2)$$

and

this may be written more compactly

$$\text{Distortion} = Au^3 + \left(B + p_3 \frac{dA}{dp_1}\right)u^5. \qquad \dots\dots(12.3)$$

As might be expected, the term in p_3 will in practice often be found to be negligible compared with the other terms, in which case we have simply

$$\text{Distortion} = Au^3 + Bu^5. \qquad \dots\dots(12.4)$$

§13. COMA AND CURVATURE OF FIELD

In this paragraph we take for the sake of simplicity an object situated at infinity. This implies no restriction in principle. (For objects finitely situated we only require the additional application of equation (11.2): see (13.5)). In equations in which several rays of a pencil are involved, quantities without suffix refer to the principal ray, whilst, as before, the suffixes a and b refer to the extreme rays.

It may be easily verified that according to the definition of §3, the coma, C', is given by

$$C' = \left\{\frac{v_a{}' + v_b{}'}{2} - v'\right\} + \frac{v_a{}' - v_b{}'}{2}\left\{\frac{2\tan U' - (\tan U_a{}' + \tan U_b{}')}{\tan U_a{}' - \tan U_b{}'}\right\},$$
$$\dots\dots(13.1)$$

whilst the curvature of field, W', is given by

$$W' = \frac{v_a{}' - v_b{}'}{\tan U_a{}' - \tan U_b{}'}. \qquad \dots\dots(13.2)$$

(13.1) may therefore be written

$$C' = \{\tfrac{1}{2}(v_a{}' + v_b{}') - v'\} - W'\{\tfrac{1}{2}(\tan U_a{}' + \tan U_b{}') - \tan U'\}. \ \dots\dots(13.3)$$

In this equation, since the lowest terms of the factor of W' are of the third degree, W' need in general only be evaluated to the $(n-1)$th order if C' is required to the nth order. Thus, when $n=2$, since $\delta y_f{}'$ and $\delta u_f{}'$ are known to the first order, the calculation of $\sin U_f{}'$ and hence of $\tan U_f{}'$ to this order presents no difficulty. Therefore (omitting the suffix f), let

$$\tan U' = u_p{}'y + u_q{}'u + u_p{}'(t_0 y^3 + t_1 y^2 u + t_2 yu^2 + t_3 u^3). \qquad \dots\dots(13.4)$$

We have assumed the incident pencil to consist of parallel rays;

$$\left. \begin{aligned} \therefore \quad y_a &= y + w, \\ y_b &= y - w, \end{aligned} \right\} \dots\dots(13.5)$$

where w is the semi-aperture of the incident pencil. (13.5) applies because $y - y_a [= (L_{a1} - L_1) \sin U_1]$ is the perpendicular distance between the a ray and the principal ray; and similarly for the b ray. (The fact that, if the pencil is actually limited physically by the diaphragm, the principal ray will not in general lie precisely halfway between the extreme rays at the first surface is of no practical consequence.)

Now to the required order, by (13.2), (13.4), (13.5), and (10.2),

$$W' = \frac{1}{u_p}\{(3a_0 p_1{}^2 + 2a_1 p_1 + a_2)u^2 + a_0 w^2\} \qquad \ldots\ldots(13.6)$$

and

$$\tfrac{1}{2}(\tan U_a' + \tan U_b') - \tan U' = u_p'(3t_0 p_1 + t_1)uw^2. \qquad \ldots\ldots(13.7)$$

Also, by (11.1) and (10.4),

$$\tfrac{1}{2}(v_a' + v_b') - v' = (3a_0 p_1 + a_1)uw^2 + \{(10b_0 p_1{}^3 + 6b_1 p_1{}^2 + 3b_2 p_1 + b_3)$$
$$+ 3a_0 p_3\}u^3 w^2 + (5b_0 p_1 + b_1)uw^4. \qquad \ldots\ldots(13.8)$$

$$\therefore \quad C' = (3a_0 p_1 + a_1)uw^2 + \{(10b_0 p_1{}^3 + 6b_1 p_1{}^2 + 3b_2 p_1 + b_3)$$
$$+ 3a_0 p_3 - (3a_0 p_1{}^2 + 2a_1 p_1 + a_2)(3t_0 p_1 + t_1)\}u^3 w^2$$
$$+ \{(5b_0 p_1 + b_1) - a_0(3t_0 p_1 + t_1)\}uw^4. \qquad \ldots\ldots(13.9)$$

As in the case of distortion, it will frequently be found in practice that the terms in p_3, t_0, t_1 may be neglected. If so, (13.9) may be written, in terms of (12.2),

$$C' = A_2 uw^2 + (B_2 u^3 w^2 + B_4 uw^4), \qquad \ldots\ldots(13.10)$$

where

and

$$\left.\begin{array}{l} A_n = \dfrac{1}{n!}\cdot\dfrac{d^n A}{dp^n}, \\[2ex] B_n = \dfrac{1}{n!}\dfrac{d^n B}{dp^n}. \end{array}\right\} \qquad \ldots\ldots(13.11)$$

By means of (13.2), a formula analogous to (13.9) may be written down, giving W' correct to the required order. From our present point of view this is of no great interest (see end of §3). Accordingly we content ourselves with observing that if the terms containing p_3, t_0, t_1, t_2 can again be neglected, the equation for W' analogous to (13.10) becomes

$$u_p'W' = (A_1 u^2 + A_3 w^2) + (B_1 u^4 + B_3 u^2 w^2 + B_5 w^4). \qquad \ldots\ldots(13.12)$$

§14. PRINCIPLES OF AN ALTERNATIVE METHOD

It is interesting from a theoretical point of view that whereas previously we made use of the *two* quasi-invariants $\Delta\Lambda$ and $\Delta\Lambda^*$, each of which splits up into two (quasi-invariant) parts, i.e.

$$\left.\begin{array}{l} \Delta\Lambda_p = c_p \Delta, \\[1ex] \Delta\Lambda_q = c_q \Delta, \end{array}\right\} \qquad \ldots\ldots(14.1)$$

and

$$\left.\begin{array}{l} \Delta\Lambda_p^* = c_p(\sec U'\Delta + \sin U\Delta(\sec U)) - c_p^* \sin I\Delta(\sec U), \\[1ex] \Delta\Lambda_q^* = c_q(\sec U'\Delta + \sin U\Delta(\sec U)) - c_q^* \sin I\Delta(\sec U), \end{array}\right\} \qquad \ldots\ldots(14.2)$$

where $\qquad c^* = Nru \qquad \ldots\ldots(14.3)$

(together with their corresponding sums D_p, D_q and D_p^*, D_q^* respectively), we can also proceed in a way which avoids the use of the former altogether.

If we now take the associated ray of the arbitrary ray R to be $(L_1 \tan U_1, \tan U_1)$ we can write down a set of equations, by (5.1), similar to (4.13), but with the sines of the angles replaced by their tangents, and with the $\Delta\Lambda$ replaced by the $\Delta\Lambda^*$. Write

$$\left.\begin{aligned}\delta y_j &= -\frac{1}{N_1}\sum_{\nu=1}^{j-1}\Delta\Lambda_{q\nu}^*,\\[4pt]\delta u_j &= +\frac{1}{N_1}\sum_{\nu=1}^{j-1}\Delta\Lambda_{p\nu}^*,\end{aligned}\right\}\quad\ldots\ldots(14.4)$$

and similarly $\delta y_j^{*\prime}$ and $\delta u_j^{*\prime}$, as before. Then in place of (5.9) we have now

$$\left.\begin{aligned}\tan U_j &= u_{pj}(y+\delta y_j^*)+u_{qj}(u+\delta u_j^*),\\\ L_q\tan U_j &= y_{pj}(y+\delta y_j^*)+y_{qj}(u+\delta u_j^*),\\\ \sec U_j\sin I_j &= i_{pj}(y+\delta y_j^*)+i_{qj}(u+\delta u_j^*),\\\ &\qquad\qquad\ldots\ldots\ldots\ldots\ldots\text{, etc.}\end{aligned}\right\}\quad\ldots\ldots(14.5)$$

$\Delta\Lambda_p^*$ and $\Delta\Lambda_q^*$ are now to be regarded as expanded in series of ascending powers of $\tan U$. If, as before, we absorb, in the nth term of these series, $\tan^{2n+1} U$ in the factor that multiplies it, then the nth term will be composed solely of the products of quantities given by (14.5). Hence we may proceed as before.

Although $\Delta\Lambda$ is no longer explicitly contained in these equations, it is doubtful whether their practical application would be more convenient than that of the formulae developed earlier.

§15. EXAMPLE OF COMPUTATIONS

(a) The system used for the purpose of illustrating the practicability of the methods above is a slightly altered design of a corrected Cooke Triplet as given by Hardy and Perrin (1932, Chap. 21). The specifications of the system are

Surface	1	2	3	4	5	6	
r	$+20 \cdot 05$	$-128 \cdot 3$	$-58 \cdot 8$	$+18 \cdot 9$	$+311 \cdot 3$	$-66 \cdot 2$	
N'	$1 \cdot 6162$	1	$1 \cdot 5725$	1	$1 \cdot 6162$	1	$\ldots\ldots(15.1)$
d'	$3 \cdot 896$	$1 \cdot 63$	$0 \cdot 93$	$13 \cdot 42$	$3 \cdot 03$	–	

$$(f' = 96 \cdot 729);\ d = 4 \cdot 0 \text{ (behind 4th surface)}$$
$$\text{Max. stop-number of system} = f'/5 \cdot 6.$$

The object is assumed to be situated at infinity, so that we are essentially computing v_p'. Moreover the factor $N_1'u'_{0f}$ may be replaced by $1/f'$ wherever it occurs where f' is the focal length of the system.

(b) *Table* 1

This gives the complete detailed computation for v' (primary and secondary). Little comment is needed, as the computing scheme is arranged so as to explain itself. (A lowered number after a nought following the decimal point gives the total number of noughts following the latter.) A current symbol, t_ν, is attached to such quantities as recur at a later stage of the computation. The entries are

Table 1

$j\rightarrow$		1	2	3	4	5	6	$f'\Sigma f$
l_p		∞	48·6921	24·6892	50·1798	942·331	528·503	
u_p		0	0·0190154	0·0351802	0·0169942	0·$_2$892248	0·00158183	
$(l_p-r)/r=x_p$			-1·37952	-1·41989	1·65502	2·02708	-8·98343	
$x_p u_p=i_p$		0·0498753	-0·0262325	-0·0499518	0·0281257	0·00180866	-0·0142102	
$Ni_p/N'=i_p'$		0·0308596	-0·0423970	-0·0317658	0·0442277	0·00111908	-0·0229666	
$(i+u-i')_p=u_p'$		0·0190157	0·0351802	0·0169942	0·$_2$892248	0·00158183	0·0103382	
$i_p'/u_p=x_p'$		1·62285	-1·20514	-1·86921	49·5688	0·707462	-2·22153	
$(x_p'+1)r=l_p'$		52·5881	26·3192	51·1098	955·751	531·533	80·8654	
l_q		0	-3·8960	-4·06883	-7·08442	-17·3842	-32·1276	
u_q		1·0	0·618735	0·988422	0·653469	1·167810	0·697701	
$(l_q-r)/r=x_q$		-1·0	-0·969634	-0·930802	-1·37484	— 1·05584	— 0·514688	
$x_q u_q=i_q$		-1·0	-0·599466	-0·920026	-0·898413	— 1·23303	— 0·359099	
$Ni_q/N'=i_q'$		-0·618735	-0·969634	-0·585072	-1·41275	0·762916	0·580375	
$(i+u-i')_q=u_q'$		0·618735	0·988422	0·653469	1·167810	0·697701	0·918978	
$i_q'/u_q=x_q'$		-1·0	-0·980991	-0·895333	-1·20975	-1·09346	-0·631544	
$(x_q'+1)r=l_q'$		0	-2·43883	-6·15442	-3·96421	-29·0976	-24·3918	
$(u+i)_p$	t_1	0·0498753	-0·0072168	-0·0147716	0·0451199	0·00270091	-0·0126284	
$(i-i')_p$	t_2	0·0190157	0·0161645	-0·0181859	-0·0161020	0·$_2$689578	0·$_2$875634	
$(i'-u)_p$	t_3	0·0308596	-0·0614127	-0·0669459	0·0272335	0·$_2$226834	-0·0245484	
$\frac{1}{2}(u'-3u)_p$	t_4	0·00475392	-0·0010050	-0·00753671	-0·0149809	-0·$_3$486167	0·$_3$919616	
$Nri_p=c_p$		1·0	5·43953	2·93716	0·835903	0·563036	1·52039	
$(u+i)_q$	t_5	0	0·0187887	0·0683966	— 0·244944	— 0·0652151	0·338602	
$(i-i')_q$	t_6	— 0·381265	0·369687	-0·334954	0·514341	— 0·470109	0·221277	
$(i'-u)_q$	t_7	— 0·161874	-0·158837	-1·57349	— 2·06622	1·93073	-1·27808	
$\frac{1}{2}(u'-3u)_q$	t_8	— 0·595316	-0·0398832	-0·196771	— 0·237048	1·24580	-0·193064	
$Nri_q=c_q$		-20·05	124·404	54·0975	-26·7011	-383·841	38·4208	
$c_q/c_p=\theta$		-20·05	22·8704	18·4183	-31·9428	-681·734	25·2704	

p trace (rows 1–8), *q trace* (rows 9–16)

Group	Expression	t						
r_3	$u_\rho{}^2$	t_9	0	$0{\cdot}0_3361597$	$0{\cdot}00123764$	$0{\cdot}0_3288803$	$0{\cdot}0_579611$	$0{\cdot}0_525022$
	$i_\rho{}^2$		$0{\cdot}00248755$	$0{\cdot}0_3688144$	$0{\cdot}00249518$	$0{\cdot}0_3791056$	$0{\cdot}0_532713$	$0{\cdot}0_3201930$
	$i_\rho'^2$		$0{\cdot}0_3952316$	$0{\cdot}00179750$	$0{\cdot}00100907$	$0{\cdot}00195609$	$0{\cdot}0_512523$	$0{\cdot}0_3527463$
	$u_\rho'^2$	t_{10}	$0{\cdot}0_3361597$	$0{\cdot}00123764$	$0{\cdot}0_328803$	$0{\cdot}0_379611$	$0{\cdot}0_525022$	$0{\cdot}0_3106878$
	$\tfrac{3}{8}\Sigma$	t_{11}	$0{\cdot}00142555$	$0{\cdot}00153183$	$0{\cdot}00188651$	$0{\cdot}00113878$	$0{\cdot}0_529332$	$0{\cdot}0_3314540$
	$u_\rho u_q$	t_{12}	0	$0{\cdot}0117657$	$0{\cdot}0347729$	$0{\cdot}0111052$	$0{\cdot}00104198$	$0{\cdot}00110364$
	$i_\rho i_q$		$-0{\cdot}0498753$	$0{\cdot}0157381$	$0{\cdot}0459569$	$-0{\cdot}0252685$	$-0{\cdot}00223012$	$0{\cdot}00510287$
	$i_\rho' i_q'$		$-0{\cdot}0190939$	$0{\cdot}0411095$	$0{\cdot}0185853$	$-0{\cdot}0624828$	$-0{\cdot}0_3853766$	$0{\cdot}0133292$
	$u_\rho' u_q'$	t_{13}	$0{\cdot}0117657$	$0{\cdot}0347729$	$0{\cdot}0111052$	$0{\cdot}00104198$	$0{\cdot}00110364$	$0{\cdot}0_3950054$
	$\tfrac{3}{8}\Sigma$	t_{14}	$-0{\cdot}0429027$	$0{\cdot}0775396$	$0{\cdot}0828152$	$-0{\cdot}0567031$	$-0{\cdot}0_3703704$	$0{\cdot}0217772$
	$u_q{}^2$	t_{15}	$1{\cdot}0$	$0{\cdot}382833$	$0{\cdot}976979$	$0{\cdot}427021$	$1{\cdot}36378$	$0{\cdot}486787$
	$i_q{}^2$		$1{\cdot}0$	$0{\cdot}359936$	$0{\cdot}846447$	$0{\cdot}807146$	$1{\cdot}52035$	$0{\cdot}128952$
	$i_q'^2$		$0{\cdot}382833$	$0{\cdot}940189$	$0{\cdot}342309$	$1{\cdot}99588$	$0{\cdot}582041$	$0{\cdot}336835$
	$u_q'^2$	t_{16}	$0{\cdot}382833$	$0{\cdot}976979$	$0{\cdot}427021$	$1{\cdot}36378$	$0{\cdot}486787$	$0{\cdot}844520$
	$\tfrac{3}{8}\Sigma$	t_{17}	$1{\cdot}03713$	$0{\cdot}997477$	$0{\cdot}972284$	$1{\cdot}72268$	$1{\cdot}48236$	$0{\cdot}673910$
r_2	$\tfrac{1}{2}(t_{10}-t_9)$	t_{18}	$0{\cdot}0_3180798$	$0{\cdot}0_3438023$	$-0{\cdot}0_3474420$	$-0{\cdot}0_3144004$	$0{\cdot}0685303$	$0{\cdot}0_4521877$
	$(t_{13}-t_{12})$	t_{19}	$0{\cdot}0117657$	$0{\cdot}0230072$	$-0{\cdot}0236676$	$-0{\cdot}0_3100632$	$0{\cdot}0_261666$	$0{\cdot}00839690$
	$\tfrac{1}{2}(t_{16}-t_{15})$	t_{20}	$-0{\cdot}308583$	$0{\cdot}297073$	$-0{\cdot}274979$	$0{\cdot}468380$	$-0{\cdot}438497$	$0{\cdot}178866$
	$\tfrac{1}{2}c_p t_1+t_2+t_8$	t_{21}	$0{\cdot}0_4146338$	$0{\cdot}0_4194847$	$-0{\cdot}0_4264111$	$-0{\cdot}0_6826944$	$0{\cdot}0_9118943$	$0{\cdot}0_6206356$
	$\tfrac{1}{2}c_p t_5+t_6+t_7$	t_{24}	0	$-0{\cdot}0300064$	$0{\cdot}0529399$	$0{\cdot}108798$	$-0{\cdot}0166638$	$-0{\cdot}0727958$
r_1	t_5/t_1		0	$-2{\cdot}60347$	$-4{\cdot}63027$	$-5{\cdot}42874$	$-24{\cdot}1456$	$-26{\cdot}8128$
	t_6/t_2		$-20{\cdot}05$	$22{\cdot}8704$	$18{\cdot}4183$	$-31{\cdot}9428$	$-681{\cdot}734$	$25{\cdot}2704$
	t_7/t_3		$-52{\cdot}4548$	$25{\cdot}8639$	$23{\cdot}5039$	$-75{\cdot}8707$	$-8511{\cdot}63$	$52{\cdot}0636$
	Σ	t_{25}	$-72{\cdot}5048$	$46{\cdot}1308$	$37{\cdot}2919$	$-113{\cdot}242$	$-9217{\cdot}51$	$50{\cdot}5212$

Table 1—continued

j →	1	2	3	4	5	6	f'Σf₁	
γ₁ $\;t_1/t_5$	—	−0.384102	−0.215970	−0.184205	−0.0414154	−0.0372956		
t_2/t_6	—	0.0437247	0.0542940	−0.0313060	−0.0014668	0.0395719		
t_3/t_7	—	0.0386640	0.0425461	−0.0131803	-0.0_311750	0.0192073		
Σf $\;t_{26}$	—	−0.301714	−0.119130	−0.228691	−0.0429997	0.0214836		
$t_{21}t_{25}\;\;t_{22}$	−0.00106102	$0.0_3898846$	$-0.0_3984922$	$0.0_3936450$	$-0.0_5109633$	$0.0_3104253$		
$t_{24}t_{26}\;\;t_{23}$	0.0153907	0.00905335	−0.00630673	−0.0248812	$0.0_3716538$	−0.00156392		
γ₁+γ₂ $\;t_{21}\;\;t_{31}$	$-0.0_3880226$	0.00133687	−0.00145934	$0.0_3792446$	-0.0_624330	$0.0_3156441$	$0.0_3145259$	a_0
$t_{22}+t_{18}\;\;t_{37}$	0.0271564	0.0320605	−0.0299744	−0.0349444	$0.0_3778204$	0.00683298	−0.0052286	a_1
$t_{23}+t_{19}\;\;t_{38}$	−0.308583	0.267066	−0.222039	0.577178	−0.455161	0.106071	0.18468	a_2
$t_{24}+t_{20}\;\;t_{39}$							−3.4308	a_3
$\theta t_{41}\;\;(t_{48})$	$-0.0_5293409$	$0.0_5445623$	$-0.0_5486447$	$0.0_5264149$	-0.0_58106	$0.0_5521471$		
θt_{42}	0.0212735	0.0205569	−0.0181406	−0.0299128	$0.0_5747406$	0.00263453		
θt_{43}	−0.308583	0.207053	−0.116159	0.794774	−0.488488	−0.0395209		
θt_{44}	0	−0.686258	0.975061	−3.47532	11.3603	−1.83958		
δu $\;\Sigma f^{3-1}t_{31}\;\;t_{30}$	0	$0.0_4146338$	$0.0_4341186$	$0.0_5770747$	$-0.0_6561971$	$-0.0_6561852$		
$\Sigma f^{3-1}t_{32}\;\;t_{31}$	0	−0.00106102	$-0.0_3162178$	−0.00114710	$-0.0_3210650$	$-0.0_3211746$		
$\Sigma f^{3-1}t_{33}\;\;t_{32}$	0	0.0153907	0.0244440	0.0181373	−0.00674389	−0.00602735		
$\Sigma f^{3-1}t_{34}\;\;t_{33}$	0	0	−0.0300064	0.0229335	0.131732	0.115068		
−δy $\;\Sigma f^{3-1}\theta t_{31}\;\;t_{34}$	0	$-0.0_5293409$	$0.0_4152215$	$-0.0_5334237$	$-0.0_7708388$	$-0.0_7701649$		
$\Sigma f^{3-1}\theta t_{32}\;\;t_{35}$	0	0.0212735	0.0418305	0.0236900	−0.00622291	−0.00547550		
$\Sigma f^{3-1}\theta t_{33}\;\;t_{36}$	0	−0.308583	−0.101530	−0.217689	0.577085	0.0885970		
$\Sigma f^{3-1}\theta t_{34}\;\;t_{37}$	0	0	−0.686258	0.288804	−3.18651	8.17376		
$t_{33}t_{11}$	$0.0_4208612$	$0.0_4298473$	−0.0₄498249	$-0.0_5941706$	$0.0_{15}3$	$0.0_9649072$		
$t_{30}t_{27}$	0	$0.0_4195635$	$-0.0_4497907$	$0.0_6610776$	$0.0_{11}13673$	$-0.0_{10}878968$		
$-3t_{34}t_{31}\;\;(t_{48})$	0	$0.0_4171510$	$0.0_7120605$	$-0.0_8829185$	$0.0_{13}250$	$0.0_9434369$		

Table (coefficients b_0, b_1, b_2 with summation rows t_{38}, t_{39}, t_{40}).

Section $b_0 = -0\cdot0_5103856$ ($\Sigma\ t_{38}$)

	$0\cdot0_7208612$	$0\cdot0_7665614$	$-0\cdot0_4875551$	$-0\cdot0_7116012$	$0\cdot0_{11}13926$	$0\cdot0_9995544$
$t_{37}t_{11}$	$-0\cdot0_6125480$	$0\cdot0_5204786$	$-0\cdot0_5275306$	$0\cdot0_6902421$	$-0\cdot0_{12}7$	$0\cdot0_7492070$
$t_{31}t_{14}$	$-0\cdot0_6627831$	$0\cdot0_5151084$	$-0\cdot0_3218724$	$0\cdot0_6468903$	$-0\cdot0_{13}8$	$0\cdot0_7449386$
$t_{31}t_4$	$0\cdot0_7695682$	$-0\cdot0_6195822$	$0\cdot0_6199053$	$0\cdot0_6123884$	$-0\cdot0_{13}578$	$0\cdot0_8189768$
$t_{31}t_{27}$	0	$-0\cdot0_8141845$	$0\cdot0_6236673$	$-0\cdot0_6909015$	$0\cdot0_{10}51251$	$-0\cdot0_7331258$
$2t_{30}t_{38}$	0	$0\cdot0_6938337$	$-0\cdot0_6204537$	$-0\cdot0_5538666$	$-0\cdot0_9874656$	$-0\cdot0_8767825$
$-3t_{35}t_{31}$	0	$-0\cdot0_6124353$	$0\cdot0_5331437$	$0\cdot0_5587707$	$0\cdot0_{11}222$	$-0\cdot0_7338971$
$-2t_{34}t_{27}$	0	$0\cdot0_6784498$	$0\cdot0_6444266$	$0\cdot0_6529729$	$-0\cdot0_{10}341$	$0\cdot0_7219534$

Section $b_1 = -0\cdot0_470536$ ($\Sigma\ t_{39}$)

	$-0\cdot0_6181306$	$0\cdot0_5259998$	$-0\cdot0_4279131$	$0\cdot0_5116496$	$-0\cdot0_98559$	$0\cdot0_6111090$
$t_{38}t_{11}$	$0\cdot0_4387127$	$0\cdot0_4491113$	$-0\cdot0_4565470$	$-0\cdot0_4397939$	$0\cdot0_8228$	$0\cdot0_6214924$
$t_{37}t_{14}$	$0\cdot0_4377$	$0\cdot0_3103660$	$-0\cdot0_3120856$	$-0\cdot0_4449342$	$0\cdot0_17$	$0\cdot0_6340685$
$t_{31}t_{17}$	$0\cdot0_4151771$	$0\cdot0_4194356$	$-0\cdot0_4256791$	$-0\cdot0_4142456$	$0\cdot0_18$	$0\cdot0_6139065$
$t_{28}t_4$	$-0\cdot0_5504403$	$-0\cdot0_6903341$	$0\cdot0_5742307$	$-0\cdot0_4140289$	$0\cdot0_95330$	$-0\cdot0_6958732$
$t_{31}t_8$	$-0\cdot0_5871176$	$-0\cdot0_7777114$	$0\cdot0_5519695$	$0\cdot0_5196025$	$-0\cdot0_91481$	$-0\cdot0_6398399$
$t_{38}t_{27}$	0	$0\cdot0_4205753$	$-0\cdot0_4356722$	$0\cdot0_4143728$	$0\cdot0_816408$	$-0\cdot0_6942927$
$2t_{31}t_{38}$	0	$-0\cdot0_4680340$	$0\cdot0_5972237$	$-0\cdot0_4801694$	$-0\cdot0_6327857$	$-0\cdot0_5289371$
$3t_{30}t_{40}$	0	$0\cdot0_4117246$	$-0\cdot0_4227270$	$0\cdot0_4133457$	$0\cdot0_6767361$	$-0\cdot0_6178788$
$-3t_{36}t_{31}$	0	$0\cdot0_4180380$	$0\cdot0_5804455$	$-0\cdot0_5540050$	$-0\cdot0_9206$	$-0\cdot0_6548476$
$-2t_{35}t_{27}$	0	$-0\cdot0_5568799$	$0\cdot0_3122090$	$-0\cdot0_4375460$	$-0\cdot0_83028$	$0\cdot0_6171319$
$-t_{34}t_{48}$	0	$0\cdot0_5940683$	$0\cdot0_6456254$	$-0\cdot0_4116797$	$-0\cdot0_754540$	$0\cdot0_6479436$

Section $b_2 = 0\cdot0_393953$ ($\Sigma\ t_{40}$)

	$0\cdot0_4778980$	$0\cdot0_3105358$	$-0\cdot0_3120531$	$-0\cdot0_4577806$	$0\cdot0_449547$	$0\cdot0_5427293$
$t_{39}t_{11}$	$-0\cdot0_343990$	$0\cdot0_3409101$	$-0\cdot0_4418879$	$0\cdot0_3657278$	$-0\cdot0_513351$	$0\cdot0_5333635$
$t_{38}t_{14}$	$-0\cdot00116508$	$0\cdot00248596$	$-0\cdot00248234$	$0\cdot00198146$	$-0\cdot0_55476$	$0\cdot0_3148803$
$t_{37}t_{17}$	$-0\cdot0_3912904$	$0\cdot00133350$	$-0\cdot00141889$	$0\cdot00136513$	$-0\cdot0_63607$	$0\cdot0_3105427$
$t_{28}t_4$	$0\cdot0_4731662$	$-0\cdot0_5909863$	$0\cdot0_4475320$	$0\cdot0_3372743$	$-0\cdot0_634836$	$-0\cdot0_5143820$
$t_{28}t_8$	$0\cdot0_5631645$	$-0\cdot0_4358489$	$0\cdot0_3193804$	$-0\cdot0_3221983$	$0\cdot0_6136581$	$-0\cdot0_4201276$
$t_{38}t_{27}$	0	0	$0\cdot0_4437896$	$0\cdot0_4181735$	$-0\cdot0_532050$	$-0\cdot0_4180029$
$2t_{32}t_{38}$	0	$0\cdot0_5986867$	$-0\cdot00146539$	$-0\cdot00126760$	$-0\cdot0_4104963$	$-0\cdot0_4823696$
$3t_{31}t_{39}$	0	$-0\cdot0_3850091$	$0\cdot0_3108029$	$-0\cdot00198624$	$0\cdot0_3287638$	$-0\cdot0_4673801$
$-3t_{37}t_{31}$	0	0	$-0\cdot0_4543745$	$0\cdot0_5716474$	$0\cdot0_4114$	$-0\cdot0_4506012$
$-2t_{36}t_{27}$	0	$0\cdot0_3825071$	$-0\cdot0_3296334$	$0\cdot0_3345014$	$0\cdot0_42808$	$-0\cdot0_4277204$
$-t_{35}t_{38}$	0	$-0\cdot0_6682041$	$0\cdot00125384$	$0\cdot0_3827830$	$0\cdot0_448427$	$0\cdot0_4374140$

Table 1—continued

j →		1	2	3	4	5	6	$f'\Sigma f'_i$	
Σ	t_{41}	$-0\cdot00181307$	$0\cdot00446342$	$-0\cdot00448921$	$0\cdot00209897$	$0\cdot0_3281009$	$0\cdot0_4933736$	$0\cdot061374$	b_3
	$t_{28}t_{14}$	$0\cdot0132390$	$0\cdot0207082$	$-0\cdot0183882$	$-0\cdot0327278$	$0\cdot0_3320298$	$0\cdot00230992$		
	$t_{25}t_{17}$	$0\cdot0281646$	$0\cdot0319796$	$-0\cdot0291436$	$-0\cdot0601982$	$0\cdot00115358$	$0\cdot00460482$		
	$t_{24}t_4$	0	$0\cdot0_4301565$	$-0\cdot3_3398993$	$-0\cdot00162990$	$0\cdot0_5810138$	$-0\cdot0_4669442$		
	$t_{23}t_6$	$-0\cdot00916233$	$-0\cdot0_3361077$	$0\cdot00124098$	$0\cdot00589803$	$-0\cdot0_5892665$	$0\cdot0_3301936$		
	$2t_{28}t_{28}$	0	0	$0\cdot00179885$	$-0\cdot0_2160279$	$0\cdot0_3205028$	$0\cdot00157264$		
	$3t_{29}t_{29}$	0	$0\cdot0123310$	$-0\cdot0162826$	$0\cdot0314054$	$0\cdot0_3920866$	$-0\cdot00191798$		
	$-2t_{27}t_{27}$	0	0	$-0\cdot00200297$	$0\cdot0_3457723$	$-0\cdot0_51551$	$-0\cdot00255743$		
	$-t_{28}t_{28}$	0	$0\cdot00989334$	$-0\cdot00304330$	$-0\cdot0_2760701$	$-0\cdot0_3449090$	$-0\cdot0_2605381$		
Σ	t_{42}	$0\cdot0322413$	$0\cdot0745813$	$-0\cdot0662198$	$-0\cdot0669200$	$0\cdot0955236$	$0\cdot00364158$	$-1\cdot2694$	b_4
	$t_{29}t_{17}$	$-0\cdot320039$	$0\cdot266392$	$-0\cdot215885$	$0\cdot994295$	$-0\cdot674712$	$0\cdot714821$		
	$t_{24}t_8$	0	$0\cdot00119675$	$-0\cdot0104170$	$-0\cdot0257904$	$0\cdot0207598$	$0\cdot0140542$		
	$3t_{33}t_{30}$	-0	0	$0\cdot0199878$	$0\cdot0397101$	$-0\cdot179877$	$0\cdot0366189$		
	$-t_{37}t_{38}$	0	0	$-0\cdot0205702$	$0\cdot0100921$	$0\cdot0_2247976$	$-0\cdot0558512$	*	
Σ	t_{43}	$-0\cdot320039$	$0\cdot267589$	$-0\cdot226883$	$1\cdot018306$	$-0\cdot831349$	$0\cdot0663041$	$-2\cdot5221$	b_5

Table 2

	j →		1	2	3	4	5	6	$f'\Sigma f'_i$
$c_0 G_3$	$t_{11} - \tfrac{3}{4}t_{10}$ $t_{14} - t_{18}$	t_{44} t_{45}	$0\cdot00124475$ $-0\cdot0546684$	$0\cdot0_3910301$ $0\cdot0427668$	$0\cdot00174211$ $0\cdot0717100$	$0\cdot00113838$ $-0\cdot0577451$	$0\cdot0_51682$ $-0\cdot00180735$	$0\cdot0_3261101$ $0\cdot0122767$	
	$t_{21}t_{44}$	t_{46}	$0\cdot0_7182155$	$0\cdot0_7177898$	$-0\cdot0_7460110$	$-0\cdot0_8941376$	$0\cdot0_{15}200$	$0\cdot0_953880$	
	$t_{21}t_{44}$ $t_{21}t_{45}$		$-0\cdot0_5132071$ $-0\cdot0_6800008$	$0\cdot0_6820656$ $0\cdot0_6833299$	$-0\cdot0_5171584$ $-0\cdot0_5189394$	$0\cdot0_5106604$ $0\cdot0_6477520$	$-0\cdot0_{11}1844$ $-0\cdot0_{12}215$	$0\cdot0_7272207$ $0\cdot0_7253336$	
	Σ	t_{47}	$-0\cdot0_6212072$	$0\cdot0_6165395$	$-0\cdot0_6360978$	$0\cdot0_6154356$	$-0\cdot0_{11}2059$	$0\cdot0_6525543$	

									ord.
G_3	$\frac{3}{3}t_{46}t_9$	t_{48}	0	$0\cdot0_{11}321637$	$-0\cdot0_{10}284725$	$-0\cdot0_{11}135937$	$0\cdot0_{33}7961$	$0\cdot0_{11}674093$	
	$i_9'(i_9')^2$	t_{49}	$0\cdot0_4293881$	$-0\cdot0_7762086$	$-0\cdot0_7320539$	$0\cdot0_4865134$	$0\cdot0_8140142$	$-0\cdot0_4121140$	
	$i_9'(i_9'')$	t_{50}	$0\cdot0_8124067$	$-0\cdot0_4180517$	$-0\cdot0_3124639$	$0\cdot0_4222490$	$0\cdot0_8591667$	$-0\cdot0_5286947$	
	$u_{29}(t_{49}-t_{50})$	t_{51}	0	$-0\cdot0_1110589$	$0\cdot0_6225716$	$0\cdot0_5109212$	$-0\cdot0_{11}402872$	$-0\cdot0_7146232$	
	$-i_9i_9't_9$	t_{52}	0	$-0\cdot0_6402161$	$-0\cdot0_6196384$	$-0\cdot0_6359254$	$-0\cdot0_{11}161134$	$-0\cdot0_9816618$	
	$(i_9'')^2t_{50}$	t_{53}	$0\cdot0_6308623$	$-0\cdot0_7124222$	$-0\cdot0_6310997$	$0\cdot0_7176002$	$0\cdot0_{13}193552$	$-0\cdot0_9579432$	
	$(i_9')^2t_{40}$	t_{54}	$0\cdot0_7279868$	$-0\cdot0_6136985$	$-0\cdot0_7323446$	$0\cdot0_6169228$	$0\cdot0_{14}175500$	$-0\cdot0_8638969$	
	$(t_{63}-t_{64})/t_9$	t_{55}	$0\cdot0_4147581$	$0\cdot0_5770596$	$0\cdot0_4153224$	$0\cdot0_5941672$	$0\cdot0_{10}255231$	$0\cdot0_6663549$	
	$\frac{1}{8}(t_{61}+t_{63}+t_{65})$	t_{56}	$0\cdot0_5184476$	$0\cdot0_6774739$	$0\cdot0_5207698$	$0\cdot0_5126870$	$0\cdot0_{11}248539$	$0\cdot0_6810136$	
	$t_{31}t_{56}+t_{48}$	t_{57}	$0\cdot0_{10}269959$	$0\cdot0_{10}183119$	$-0\cdot0_{10}833274$	$-0\cdot0_{10}118509$	$0\cdot0_{21}3751$	$0\cdot0_{13}167850$	
δ	$t_{32}t_{30}$	t_{58}	0	$0\cdot0_7131536$	$-0\cdot0_7336041$	$0\cdot0_8721766$	$0\cdot0_{12}6159$	$-0\cdot0_{10}58575$	
		t_{59}							$-0\cdot0_5510463$	**2nd ord.**
	$t_{66}+t_{68}+t_{69}$	t_{60}	$0\cdot0_7182155$	$0\cdot0_4480944$	$-0\cdot0_7675546$	$-0\cdot0_7104879$	$0\cdot0_{10}6411$	$0\cdot0_991459$	
	θt_{60}	t_{61}	$-0\cdot0_6365221$	$0\cdot0_6109994$	$-0\cdot0_6124424$	$0\cdot0_6335015$	$0\cdot0_94372$	$0\cdot0_723112$	
	$\Sigma_1^i-t_{60}$	t_{62}	0	$0\cdot0_7182155$	$0\cdot0_7663099$	$-0\cdot0_812447$	$-0\cdot0_7117326$	$-0\cdot0_7117319$	
	$\Sigma_1^i-t_{61}$	t_{63}	0	$-0\cdot0_6365221$	$0\cdot0_6734714$	$-0\cdot0_6509529$	$-0\cdot0_6174514$	$-0\cdot0_6174951$	
	$(t_{64})^2$	t_{64}	0	$0\cdot0_686089$	$0\cdot0_823169$	$0\cdot0_6111714$	$0\cdot0_849118$	$0\cdot0_849231$	
	$-t_{30}t_{34}$	t_{65}	0	$0\cdot0_842937$	$-0\cdot0_851933$	$0\cdot0_825761$	$-0\cdot0_{10}39385$	$-0\cdot0_{10}39421$	
	$(t_{30})^2$	t_{66}	0	$0\cdot0_921415$	$0\cdot0_811641$	$0\cdot0_{10}59405$	$0\cdot0_{12}3158$	$0\cdot0_{12}3157$	
	$-3t_{31}t_{63}$		0	$0\cdot0_{10}21349$	$0\cdot0_{10}58214$	$-0\cdot0_{10}12641$	$0\cdot0_{10}60$	$0\cdot0_{11}10831$	
	$t_{32}t_{62}$		0	$-0\cdot0_{10}163729$	$-0\cdot0_{10}653101$	$-0\cdot0_{11}11656$	$0\cdot0_{13}1286$	$-0\cdot0_{11}112231$	
	$3t_{31}t_{64}$		0	$0\cdot0_{11}503322$	$-0\cdot0_{11}18357$	$-0\cdot0_{12}27714$	$0\cdot0_{17}18$	$0\cdot0_{13}304477$	
	$2t_{31}t_{65}$		0	$0\cdot0_{11}777188$	$0\cdot0_{12}102298$	$0\cdot0_{12}48248$	$0\cdot0_{15}86$	$-0\cdot0_{14}8220$	
	$t_{33}t_{66}$		0	$0\cdot0_{11}119388$	$-0\cdot0_{11}173416$	$-0\cdot0_{11}147807$	$0\cdot0_{15}226$	$-0\cdot0_{15}4937$	
	$-5t_{46}t_{34}$		0	$0\cdot0_{10}26098$	$0\cdot0_{10}35018$	$-0\cdot0_{10}15732$	$0\cdot0_{19}70$	$0\cdot0_{12}18902$	
	$t_{47}t_{30}$		0	$0\cdot0_{10}24204$	$-0\cdot0_9123161$	$0\cdot0_{10}11897$	$0\cdot0_{17}11$	$-0\cdot0_{12}2950$	
	t_{67}		$0\cdot0_{10}269959$	$0\cdot0_{10}18312$	$-0\cdot0_{10}833274$	$-0\cdot0_{10}118509$	$0\cdot0_{21}3751$	$0\cdot0_{13}167850$	
	Σ	t_{67}	$0\cdot0_{10}269959$	$0\cdot0_9121026$	$-0\cdot0_9177514$	$-0\cdot0_{10}28917$	$0\cdot0_{13}1324$	$0\cdot0_{12}20920$	$-0\cdot0_856284$ **3rd ord.**

reduced to six significant figures (the original calculations were generally carried out to seven). In the case of a corrected system it is important that the number of significant figures carried should exceed by at least two the number of significant figures desired in the final result. Since no trigonometrical tables are required, this creates no practical difficulty. It may, however, be convenient at times to "modify" the dimensions of the system by dividing all the given radii and distances by the focal length (or a similar convenient quantity). This avoids the continual occurrence of a great number of noughts after the decimal point.

Altogether it will be seen that the computations can be rapidly carried out if a high-speed calculating machine is available. For each surface there are 113 entries consisting entirely of simple additions and multiplications. Many of these can be further speeded up by observing that some of the coefficients repeatedly appear multiplied by various other coefficients. Thus it will be seen that in the last part of table 1, t_{28}, for instance, appears 11 times, i.e. multiplied by t_{11}, t_{14}, t_{17}, t_{34}, t_{35}, t_{36}, t_{37}, $2t_{30}$, $2t_{31}$, $2t_{32}$, $2t_{33}$ respectively.

The *first-order contributions* by the different surfaces appear in the rows t_{21}, t_{27}, t_{28}, t_{29}; whilst the *second-order contributions* appear in the rows $t_{38}, \ldots, t_{43}$ (disregarding the factor f' by which they should be multiplied to give the actual contributions as defined). Hence the final result is

$$v' = +0.0_3145259y^3 - 0.0_252286y^2u + 0.18468yu^2 - 3.4308u^3$$
$$- 0.0_5103856y^5 - 0.0_470536y^4u + 0.0_393953y^3u^2$$
$$+ 0.061374y^2u^3 - 1.2694yu^4 - 2.5221u^5. \qquad \ldots\ldots(15.2)$$

(c) *Table 2*

This gives the additional computations required to determine the secondary and tertiary spherical aberrations, measured by $(l'_{pf} - L_f')\sin U_f'$ (see (11.4)). There are only 33 entries per surface; t_{60} and t_{67} are the contributions by the individual surfaces to the secondary and tertiary aberrations respectively (except again for the factor f'). The final result is

$$\text{Sph. ab.} = +0.0_314526y^3 - 0.0_510463y^5 - 0.0_856284y^7. \qquad \ldots\ldots(15.3)$$

Since
$$\delta u_f' = \left(\overset{f}{\underset{1}{\Sigma}} t_{21}\right)y^3 + \left(\overset{f}{\underset{1}{\Sigma}} t_{60}\right)y^5,$$

and
$$-\delta y_f' = \left(\overset{f}{\underset{1}{\Sigma}} t_{68}\right)y^3 + \left(\overset{f}{\underset{1}{\Sigma}} t_{61}\right)y^5,$$

we can determine $\sin U_f'$ at once, by (5.9), correct to the second order. Hence the spherical aberration measured by $(v')_{u=0}$ may be calculated. Thus

$$\sec U_f' = 1 + 0.0_453439y^2 + 0.0_720476y^4.$$
$$\therefore \quad \text{Sph. ab.} = +0.0_314526y^3 - 0.0_510386y^5 - 0.0_856813y^7. \qquad \ldots\ldots(15.4)$$

The difference between the results (15.3) and (15.4) is negligible in practice, as will generally be the case.

(d) *Figures 1–5*

These are drawn, as explained in §11, on the basis of (15.2) and (15.4), i.e. v' is plotted against y for fixed values of u. (The units correspond to $f' = 96.729$; i.e. they represent approximately percentages of the focal length.) In this way the
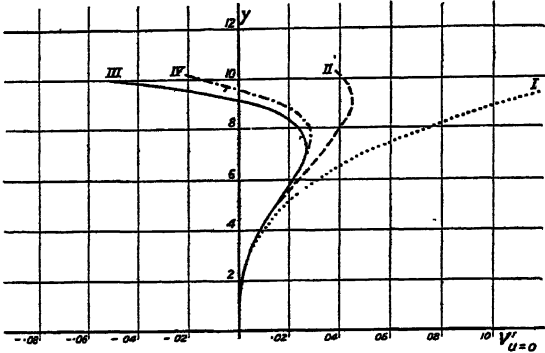
Figure 1. Sph.
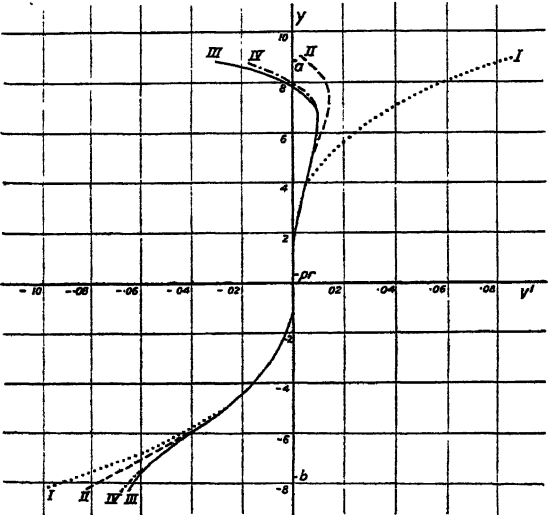
Figure 2. 3°.
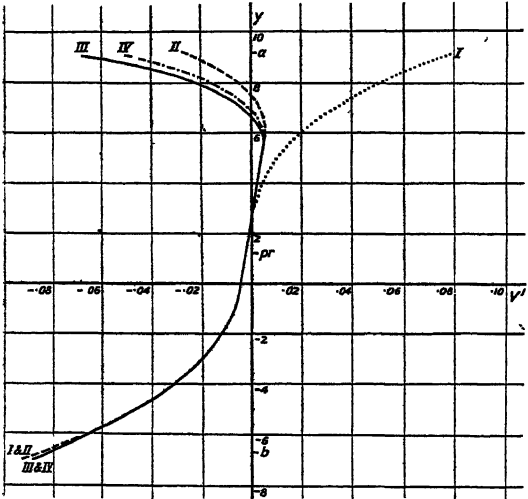
Figure 3. 6°.

values of v' corresponding to a whole pencil of rays, incident parallel at the first surface, are graphically represented for the case of pencils inclined at 0°, 3°, 6°, 12°, 18°, respectively to the axis. Roughly the whole field for which the objective is designed is thereby covered.
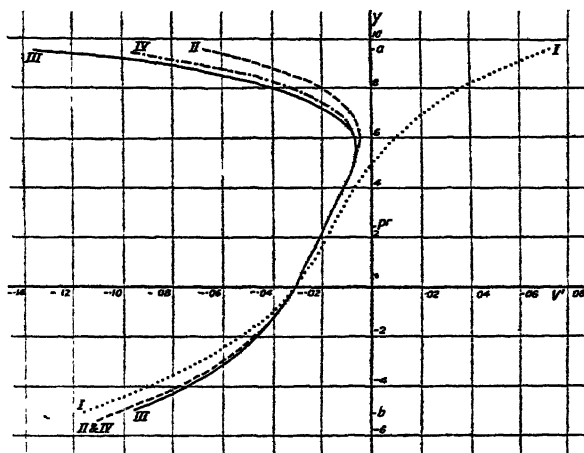
Figure 4.   12°.
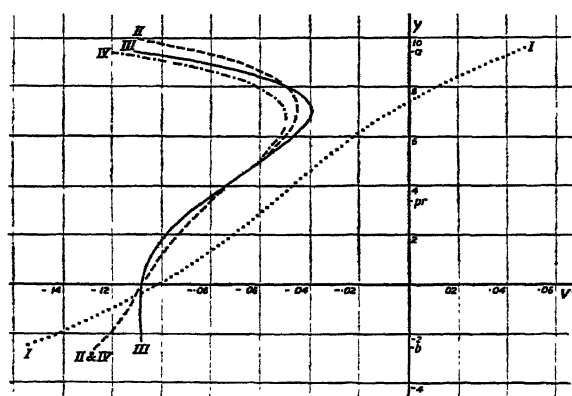
Figure 5.   18°.

The meaning of the various curves is as follows:

Curve I,   ·········, primary aberration.

Curve II,   ‑‑‑‑‑‑, (primary + secondary) aberration.   ⎫

Curve III,   ————, curve obtained from a number of   ⎬ (Figures 1–5.)
    strict trigonometrical traces.   ⎭

Curve IV,   ‑·‑··‑,   ⎰ (primary + secondary + tertiary) sph. abn.
         (Figure 1.)
    ⎱ (strict primy. + secndy.) abn. + terty. sph. abn.
         (Figures 2–5.)

The values of y marked pr, a, b refer to the principal and extreme rays corresponding to the diaphragm with the given aperture and position. (As has been pointed out, these are not involved in the calculation of v'.)

(e) Distortion and coma

The coefficients $a_0, \ldots, b_5$ being known, the calculation of distortion, coma, etc., is so simple that its inclusion here is hardly necessary. Accordingly we shall content ourselves with quoting certain results. Thus, with the given diaphragm position,

$$p_1 = 10 \cdot 9523. \qquad \ldots\ldots(15.5)$$

Putting $y = p_1 u$ in the expressions for $-\delta y_4'$, $+\delta u_4'$, (given by $(t_{30})_5, \ldots, (t_{37})_5$), we have

$$p_3 = +2 \cdot 6444. \qquad \ldots\ldots(15.6)$$

Substituting in (12.1),

$$\text{Distortion} = -1 \cdot 8444 u^3 - 8 \cdot 6839 u^5. \qquad \ldots\ldots(15.7)$$

Notice the large value of the secondary term.
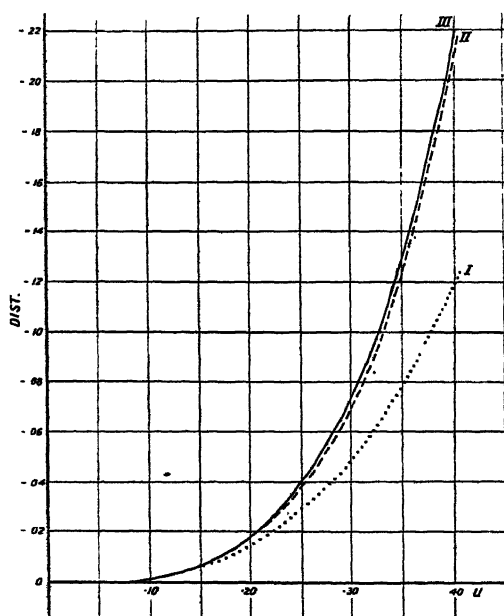
Figure 6.

As has been mentioned, the term in p_3 is often negligible in practice. Actually in this case

$$p_3 A_1 u^5 = +0 \cdot 3237 u^5. \qquad \ldots\ldots(15.8)$$

For the coma we obtain, by (13.9),

$$C' = -0 \cdot 0_3 45580 u w^2 + (0 \cdot 027238 u^3 w^2 - 0 \cdot 0_3 12949 u w^4). \qquad \ldots\ldots(15.9)$$

(If the terms in p_3, t_0, t_1 be neglected, we find

$$C' = -0 \cdot 0_3 45580 u w^2 + (0 \cdot 028985 u^3 w^2 - 0 \cdot 0_3 12741 u w^4), \qquad \ldots\ldots(15.10)$$

which, again, is nearly the same as (15.9).)

We may note here that if a pair of extreme rays is defined by a particular value of w, then the corresponding effective stop-number of the system is given by

$$\text{stop-number} = 48 \cdot 4/w. \qquad \ldots\ldots(15.11)$$

(f) *Figure* 6. This is a graphical representation of (15.7).

(g) *Figures* 7–10

In these, by means of (15.9), coma is plotted against w for various fixed values of u. (The meaning of curves I, II and III is in each case as explained in § 15 (d)).

It will be seen that from a practical point of view the primary coma gives no useful information about the residual coma in the image, whilst the inclusion of the secondary coma gives a very fair representation of the actual state of affairs. In fact, the curves strongly suggest that, in the process of correction, virtual elimination of primary coma was aimed at by the designer, leaving, however, large
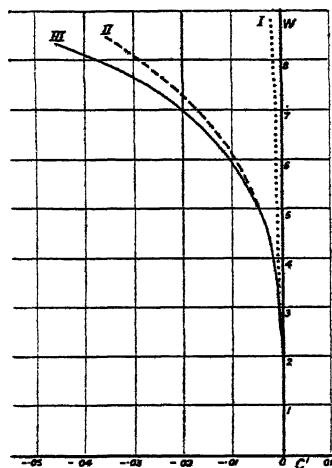
Figure 7. 3°.
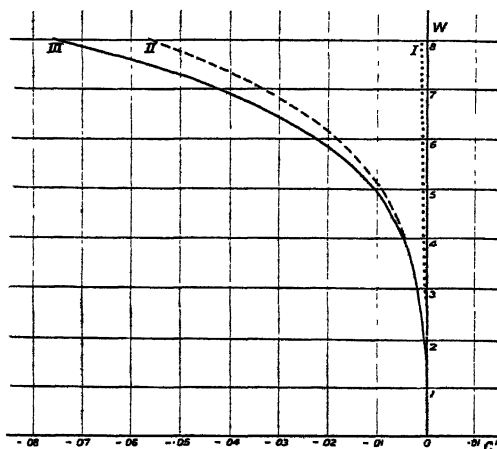
Figure 8. 6°.
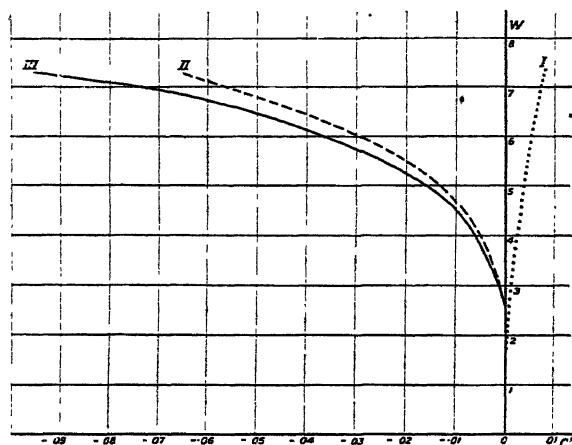
Figure 9. 12°.
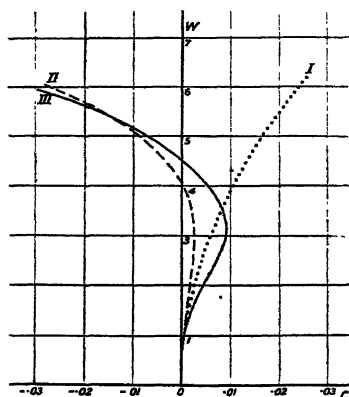
Figure 10. 18°.

secondary coma. Thus for fixed u (at any rate up to $U_1 = 12°$), the coma is virtually given by a single term proportional to the *fourth power* of the aperture of the pencil, whereas the primary term alone would predict a residual coma proportional to the *square* of the aperture.

If a direct measure of the contribution to the coma by any individual surface is required, this may be obtained from table 1 by means of a formula analogous to (13.9) or (13.10), with the as replaced by t_{21}, t_{27}, t_{28}, t_{29} and the bs replaced by $t_{38}, \ldots, t_{43}$, respectively.

§ 16. CONCLUSION

It has been shown, how, by means of relatively short and straightforward computations, the tangential aberrations of higher order than the first can be obtained for a given system. In principle, the properties of the system are completely characterized for arbitrary positions of the object and the plane of reference (and of the diaphragm) by two expressions D_p^* and D_q^*, or one such expression D^* if the object and plane of reference have fixed positions. For the calculations of D_p^* and D_q^* it is convenient, though not essential, to make use of two expressions D_p and D_q, similarly defined, the latter being needed correct only to the order next lower than that to which the D_p^* and D_q^* are required. All four expressions are sums of certain expressions, of which one is contributed by each surface of the system, and all of them vanish identically in the paraxial region.

It is hoped to extend the present analysis to cover the case of arbitrary skew rays at a later date.

REFERENCES

CONRADY, 1929. *Applied Optics and Optical Design.*
HARDY and PERRIN, 1932. *The Principles of Optics.*

ERRORS IN VISCOMETRY DUE TO SURFACE TENSION

By GUY BARR,

National Physical Laboratory, Teddington

ABSTRACT. In viscometers of the Ostwald type used for the determination of kinematic viscosities relative to water, surface tension causes a reduction of the head available and increases the time of flow. The correction is greater for water than for organic liquids, but is not proportional to the surface tension. Earlier estimates of the correction are shown to be suspect, and the classical tables of Bashforth and Adams are applied to show how the capillary rise in a measuring bulb of specified size (3·85 ml.) and form varies during the discharge. The calculated corrections for two liquids with surface tensions differing rather more than those of water and benzol diverge by 0·5% when a mean head of 10 cm. is assumed, and a simple experiment demonstrates that this estimate is probably low owing to effects of drainage on the shape of the meniscus. Two methods are proposed for eliminating the correction during the calibration of viscometers.

§ 1. INTRODUCTION

THE possibility that errors in measurement of relative viscosity may occur when liquids of different surface tension are tested in instruments of Ostwald (or similar) type has frequently been discussed (Barr, 1931) and much more frequently overlooked. The calculation of the necessary correction for viscometers having bulbs of a given shape and size should be feasible when the surface tension is known, but several attempts have been based on inadequate theory. Since the blown bulbs of a U-tube instrument are not likely to be geometrically perfect, the method used by Bénard (1907), in which the effect of

38-2

surface tension is measured at successive levels of the liquid in the bulb, would be more suitable for general use; this method is not, however, readily applicable with great accuracy to a completed viscometer, and could hardly be used with oils and opaque liquids that wet glass.

The desirability of either avoiding or correcting for the effects of surface tension has increased in recent years with the demand for higher precision in the measurement of kinematic viscosity, especially of oils. As absolute determinations have been considered impracticable for routine work, there is fairly general agreement that the viscosity of water at 20° c. be adopted as a standard, for which the value 1·007 cs. is provisionally accepted, and that all viscosities be relative to this standard. The writer is in accord with this tendency, but it must be recognized that two disadvantages attach to the selection :—

(*a*) If water gives a flow time of 200 seconds in a viscometer that may be calibrated by its means to ±0·1%, an oil with a viscosity of only 1000 cs. will require a flow time of over 55 hours. This direct comparison may, of course, be avoided by calibrating a series of viscometers by a step-up procedure, but with three instruments, allowing a maximum time of 2000 seconds for the calibrating liquids of intermediate viscosity, the cumulative errors, for two determinations of the same accuracy in each tube, will amount to $\pm \sqrt{6 \times (0 \cdot 1)^2} = \pm 0 \cdot 24\%$.

(*b*) Water has an exceptionally high surface tension (72·7 dynes/cm. at 20°), while the value for most mineral oils is about 30 dynes/cm. If the surface-tension correction with water in a viscometer amounts to only 0·5% and the correction is proportional to surface tension, the error in a comparison between water and oil will be some 0·3%; actually it will be shown later that the correction falls more rapidly, so that the error will be greater.

§2. RECENT CALCULATIONS OF AN APPROXIMATE CORRECTION FOR SURFACE TENSION

Mention may be made of three attempts later than those cited by Barr (1931) to obtain an estimate of the magnitude of the effect of surface tension on the hydrostatic head in a viscometer. In the first (Barr, 1933) the "capillary" rises, h_σ, at various levels in a bulb of the shape specified for B.S. viscometers were deduced from a table due to Sugden (1921) and were subtracted from the hydrostatic head h operating during the outflow of successive elements of volume δV. Summation of $\delta V/h$ and of $\delta V/(h-h_\sigma)$ over the whole discharge gave quantities proportional to the total time of efflux without and with correction for surface tension (σ). For bulbs of 1 ml. and 4·7 ml. capacity the corrections were calculated as 2·1% and 0·8% respectively, for water fillings. This calculation is invalidated by the omission of any attempt to allow for the effect of the slope of the walls of the 45° cones on the capillary rise.

In a long paper dealing with the use of his "suspended level" in viscometry, Ubbelohde (1936) analysed the effect of surface tension in a spherical bulb on the assumption that, when the radius at the line of contact of the meniscus with the wall is r, the capillary rise is simply $2\sigma \cos\theta/(r\rho g)$, where θ is the deviation of the wall from the vertical, g is the acceleration due to gravity, and ρ is the density of the liquid (more strictly ρ is the density of the liquid minus that of air). For the whole bulb he thus finds that the operative head is reduced by $2\sigma/(R\rho g)$, where

R is the radius of the bulb. Relying on this value, he deduces a radius for a hemispherical "suspended level" at the lower end of the capillary such that the correction should be eliminated. The calculation ignores the variations in shape of the meniscus during its passage down the bulb. The experimental comparisons of viscosities of benzol and water, determined (i) by applying the indicated correction to results obtained with a plane "suspended level", and (ii) by using the hemispherical level, differ by 0·7% (or by nearly 2% if kinetic-energy corrections are similar in the two series).

Jones and Fornwalt (1938) used a method of summation similar to that indicated by Barr (1933) for the effect of surface tension in two viscometers: one (T) with a cylindrical length of 3·8 cm., radius 0·90 cm., and conical ends, and one (F) with a cylindrical length of 5·6 cm., radius 0·85 cm., and hemispherical ends: the discharge in the first case was into a lower reservoir of radius 3·6 cm., and in the latter into one with a radius equal to that of the bulb. They assumed again that capillary rises are inversely proportional to radius, but they allowed for variations of slope of the wall by postulating that the rise was also proportional to $(1 - \sin\theta)/\cos\theta$; no explanation of this postulate is given. Since their calculations indicated corrections of 0·44% and of 0·02% for comparisons of the viscosity of water and methanol in instruments T and F, respectively, and experiment gave a difference of 0·4% in the relative times of flow, they concluded that the correction was practically eliminated by the substitution of hemispherical for conical ends in the measuring bulb and the provision of a recipient bulb of the same diameter.

§3. APPLICATION OF THE TABLES OF BASHFORTH AND ADAMS

Gross errors may be inherent in any estimation of the effect of surface tension on viscometry when it is assumed that the capillary rise is proportional to surface tension and inversely proportional to density and to radius. Thus for water ($\sigma = 72\cdot7$ dynes/cm., $\rho = 0\cdot9970$ gm./ml.) and benzol ($\sigma = 28\cdot9$ dynes/cm., $\rho = 0\cdot8774$ gm./ml.) at 20° c. the capillary rises in tubes of three different radii r cm. may be obtained from Sugden's tables as—

	$r=0\cdot2$	$r=0\cdot4$	$r=0\cdot8$	$r=1\cdot6$
Water	0·683 cm.	0·273 cm.	0·067 cm.	0·0046 cm.
Benzol	0·280 cm.	0·090 cm.	0·010 cm.	0·0002 cm.
Ratio B/W	0·41	0·33	0·15	0·04

The last row gives the ratio of the rises for benzol and water in the same tube; the ratio of the values of σ/ρ is 0·45. Now Sugden's tables were based mainly on the much more comprehensive work of Bashforth and Adams (1883), who showed how the form of the meniscus dividing a liquid from air could be calculated in detail. The contour of a meniscus is defined by the value of a parameter $\beta = b^2 \rho g / \sigma$, where b is the radius of curvature at the vertex (O in figure 1); tables (No. II) are given for values of β up to 100, showing the ratio x/b and z/b all round a sessile drop, x being the radius at a vertical distance z from the vertex. If σ and ρ are known, each value of β fixes a value for b and hence a series of values for x and z.

The points of which the co-ordinates x and z are thus listed are at intervals of 5°, from 0° to 180°, of the angle ϕ between the vertical through the axis of revolution of the drop or meniscus and the normal to the surface. Sugden's table refers to the special case $\phi = 90°$, which applies to the meniscus of a liquid in a vertical cylinder—when, as is assumed throughout this paper, the contact angle is zero—but the complete work of Bashforth and Adams allows the capillary rise to be found for any small vessel of which the internal surface constitutes a surface of revolution. The simplest form of bulb for calculation is the biconical, but the form specified for B.S. viscometers, viz., a cylinder with conical ends, may be treated by finding the shape of the meniscus at successively lower levels in the bulb when the angle ϕ is first 135° (for the upper cone), then
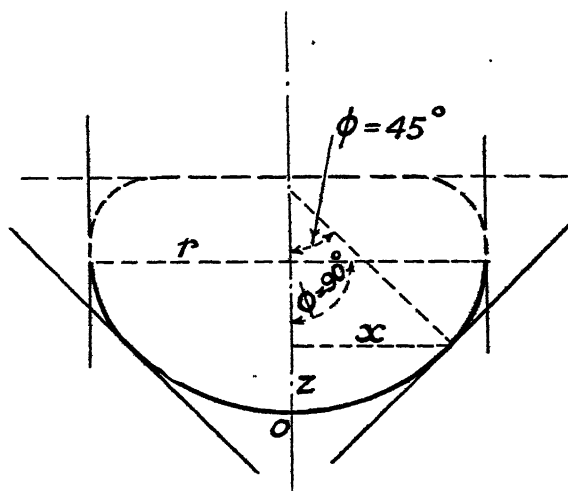
Figure 1. Diagram to illustrate definitions of x, z, r and ϕ. The contour is that of Bashforth and Adams' smallest drop (of mercury resting on the horizontal plane, shown dotted); it has been inverted so that the part below $\phi = 90°$ represents the meniscus of a liquid wetting the wall of a vertical cylinder and touching a cone, of 45° semi-vertical angle, below it.

90° (for the cylindrical part) and finally 45° (for the lower cone). For the small elements of volume enclosed between the fiducial marks on the tubing above and below the bulb, the angle ϕ is 90°, and the capillary rise may be found by either method. A constant a is defined by the relation $a^2 = 2\sigma/(\rho g)$, where ρ is again strictly the difference in density between the fluids on the inner and outer sides of the meniscus, i.e., in our case, the "apparent" density of the liquid in the viscometer. The capillary rise h_σ is given by $h_\sigma = a^2/b$; Sugden's table gives r/b for various values of r/a, where r is the radius of a vertical tube containing the liquid, while we shall use the x of Bashforth and Adams' tables to refer to the radius of the circle of contact between the meniscus and the bulb.

Method of calculation for a specified bulb

The surface-tension correction has been calculated for a bulb consisting of a cylindrical portion with a length equal to its radius, 0·896 cm., connected above and below by 45° (semi-vertical angle) cones to cylindrical tubing of radius

0·21 cm., with timing marks 0·42 cm. beyond the junctions of the tubing with the cones. The form of the bulb is nearly the same as that which was shown by Barr (1933) to give the minimum ratio of wall surface to volume for a cylinder with conical ends; the length of the cylinder is assumed equal to the radius (instead of to 0·8 of the radius) since in practice it is convenient to lengthen it even more. The actual radii result from arbitrary selection of certain round values of β (1 and 100) as minimum and maximum and of a value $a^2 = 0·16$ cm.² corresponding with a surface tension somewhat higher than that of water (for which $a^2 = 0·149$ cm.² at 20° c.). The total volume of the bulb was then subdivided by finding the volumes discharged when β assumed the values $\beta = 1, 4, 10, 40$ and 100 as the meniscus fell in the upper cone; while the cylindrical part was discharging, β remained constant at 100 and in the lower cone it fell again, to $\beta = 1·5$, before the junction with the tubing below. In the upper cone, particularly, the lowest point of the meniscus is so much below the circle of contact that a considerable addition must be made to the volume delivered from the frustrum above the plane of this circle; Bashforth and Adams provide a table (No. III) which gives the value of this addition, in terms of b, for all values of β and of ϕ. For each element of volume discharged, ΔV, the capillary rise h_σ was assumed to be the mean between those characterized by the values of β at the beginning and end of the discharge of that element. The product $\Delta V \cdot h_\sigma$ was finally summed for the whole bulb volume V and divided by V to give a mean h_σ. Assuming that the recipient bulb has the same diameter as the measuring bulb, the capillary rise in this arm will have the same value as that already calculated for the cylindrical part of the measuring bulb; the excess of h_σ over this value represents the correction to be applied to the mean head causing the flow. The method of calculation may be illustrated by detailing the procedure for the first two elements of volume considered:—

(*a*) The tubing above the upper cone has radius 0·21 cm. and the graduation is 0·42 cm. above the junction, assumed sharp. The meniscus starts with its lowest part in the plane of the graduation (see figure 2, position I) and its shape will remain unchanged until the circle of contact reaches the joint (position II). With $a = 0·40$ cm. we have $r/a = 0·525$, whence from Sugden's table $r/b = 0·9165$ and $b = 0·229$. The capillary rise in the tube is, therefore, $h_\sigma = a^2/b = 0·698$ cm. We require to know the distance of the vertex of the meniscus below the plane of the joint and also, for later use, the "bubble" volume v bounded by this plane and the surface of the meniscus. We find first the value of $\beta = 2b^2/a^2 = 0·655$. Interpolation in Bashforth and Adams' table II for $\beta = 0·655$ and $\phi = 90°$ gives $x/b = 0·917$, $z/b = 0·825$, whence $x = 0·210$ and $z = 0·189$ cm. The volume ΔV discharged while h_σ remains at 0·698 cm. is, therefore, $\pi(0·21)^2(0·42 + 0·19) = 0·082$ ml. Hence $\Delta V \cdot h_\sigma = 0·0590$. Bashforth and Adams' table III gives the volume of the "bubble" as $1·47b^3 = 0·0176$ ml.

(*b*) We pass next to the position where $\beta = 1$; here $b^2 = \frac{1}{2}a^2\beta = 0·08$ and $b = 0·282$, whence $h_\sigma = a^2/b = 0·568$ cm. For $\phi = 135°$ the tables give $r/b = 0·7369$ and $v = 2·091b^3$, whence $r = 0·21$ cm. and $v = 0·0473$ ml. The meniscus has, therefore, become tangential to the cone at the junction with the tubing above it (position III) and the volume run out since the meniscus touched the joint at 90° is $0·0473 - 0·0176 = 0·0297$ ml. This volume is considered to have been discharged while the mean capillary rise was $\frac{1}{2}(0·698 + 0·568) = 0·633$ cm.

The radius of the cylindrical part of the bulb was fixed at 0·896 cm. because this value of r results from the figure given by the table for $\beta = 100$ at $\phi = 90°$; the meniscus touches the upper cone at the same time, when $r = 0·846$ cm. for $\phi = 135°$ (position VII). In this position the lowest point of the meniscus is 0·410 cm. below the circle of contact with the cylindrical wall and 0·536 cm. below the circle of contact with the cone; a surprisingly large fraction of the total content of the bulb is, therefore, discharged before the minimum capillary rise occurs. This minimum persists until the meniscus touches the lower cone (cf. figure 1), after which there is a relatively rapid increase. The variations in capillary rise during discharge are plotted in figure 3 (upper curve).
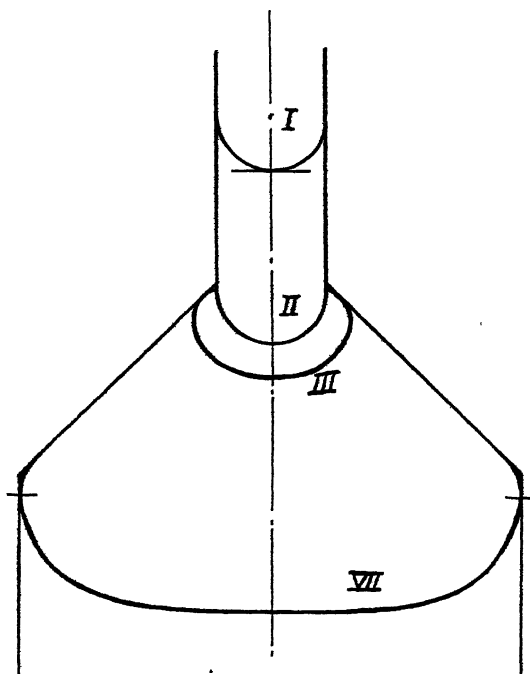
Figure 2. Positions of the meniscus at four stages of the flow from a bulb shaped as in figure 3—not to scale.

The total volume of the bulb between the timing marks, 3·85 ml., was sub-divided into 13 elements in all. The sum $\Sigma(\Delta V . h_\sigma)$ divided by 3·85 gave $h_\sigma = 0·1241$ cm. If the discharge is supposed to be into a recipient cylinder of radius 0·896 cm., in which the capillary rise is 0·0566 cm., the mean amount by which the driving head is reduced is $0·1241 - 0·0566 = 0·0675$ cm. Assuming that the average hydrostatic head is 10 cm., the surface-tension correction will therefore amount to some 0·7% when the liquid tested has the surface-tension postulated.

As Bashforth and Adams' tables do not extend beyond $\beta = 100$ it is not possible to complete a precise calculation by the above method for a liquid of lower surface tension in the postulated bulb. The positions of seven points on the lower curve of figure 3 were, however, calculated for a liquid for which $a^2 = 0·0625$ (the value of a^2 for benzol at 20° c. is 0·0671). These points refer to positions of the meniscus

in the upper and lower cones. The capillary rise in the cylindrical part of the bulb was found from Sugden's table, and the extremities of the horizontal straight line were displaced in the direction of the h_σ axis by an amount arbitrarily estimated to allow for the flattening of the meniscus; the curve was then completed by the dotted portions. Measurements, with a planimeter, of the area between this curve and the ordinate $h_\sigma = 0.0057$ and of that between the curve for $a^2 = 0.16$ and the ordinate $h_\sigma = 0.0566$ gave a ratio $0.29 : 1$. The surface-tension correction when $a^2 = 0.0625$ and the hydrostatic head is 10 cm. is, therefore, estimated as $0.29 \times 0.68\% = 0.20\%$. The correction falls more rapidly than does a^2, for which the ratio is $0.0625/0.16 = 0.39$.

The correction for water in a larger bulb of the same shape is also not directly obtainable by the use of the tables. For a smaller bulb, of volume 2·3 ml. and
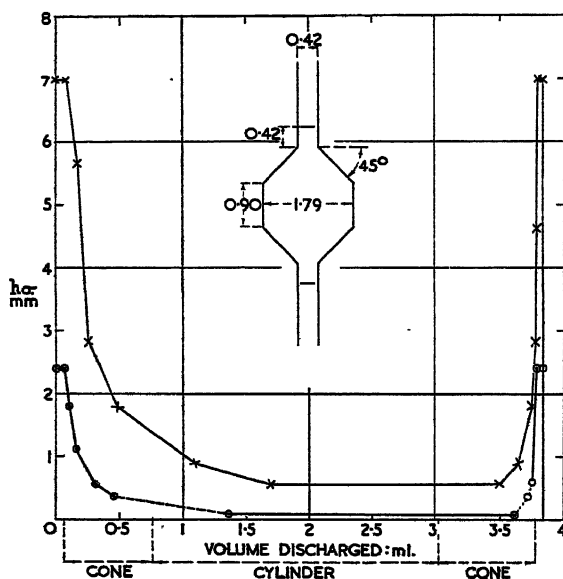
Figure 3. Calculation of surface-tension correction for viscometer bulb of volume: 3·85 ml.

maximum radius 0·75 cm., with a recipient of the same radius, the correction is found to be 0.88% (instead of 0.68%) for a hydrostatic head of 10 cm. when $a^2 = 0.16$.

Examination of the curves illustrates the correctness of the common assumption that an increase in the volume or in the maximum diameter of the bulb tends to decrease the effect of surface tension. They show, however, that an increase in the diameter of the tubing above and below the bulb will produce only minor improvement, the main portion of the effect being due to conditions in the upper cone. If adequate drainage occurs with a cone of greater semi-vertical angle than 45°, it might be preferable to adopt an angle of 60°, say, for the upper cone, so that the meniscus more rapidly reached its maximum diameter. Observation shows that some improvement in the drainage from the lower cone is desirable, and the curves suggest that the capillary effect is here so small that a steeper slope, say a 30° cone, might be adopted without undue increase in the surface-tension

correction. Such a change would require that the maximum diameter be increased and the parallel portion practically eliminated so as to prevent inconvenient increase in the overall length. A 60° cone and a 30° cone, base to base, would need to have a maximum radius 29% greater than that of the form assumed in the above calculations, in order to have the same content, and would have almost exactly the same length.

§4. CRITICISM OF THE METHOD

A stricter method of calculating the correction for surface tension requires that the height of the measuring bulb above the recipient be known, so that for each element of volume, ΔV, discharged, a quantity $\Delta V/(h - h_\sigma)$ may be deduced which is proportional to the time required for the discharge of that volume. The sum $\Sigma \Delta V/(h - h_\sigma)$ for the whole bulb is then proportional to the flow time observed in an experiment. This method, which was used by Barr (1933) and by Jones and Fornwalt (1938), takes account of the fact that the flow time is less sensitive to variations in the driving head in the earlier than in the later stages of the discharge. For a viscometer of normal design the simpler procedure now adopted will produce only a small error in the correction, and the summation of $\Delta V . h_\sigma$ or the plotting of h_σ against volume allows examination of the behaviour of different bulbs without reference to one specified mean head.

A more serious defect is common to all calculations of this type. It has been demonstrated in several papers (e.g. Jones and Stauffer, 1937) that the " drainage error ", due to variation in the volume of liquid left on the walls of the bulb, may be serious when externally applied pressures are used to shorten the flow time, but becomes negligibly small for viscometers of the type here considered, in which the times allowed for drainage are, of necessity, proportional to the viscosity of the liquid. Nevertheless, as the meniscus moves down the wall it must be distorted by the flow of liquid from the still wet surfaces above it and the distortion will be most severe precisely in those regions where it will cause the greatest effect, viz., where diameters are small.

§5. EXPERIMENT TO ILLUSTRATE THE EFFECT OF SURFACE TENSION

A simple experiment has been made to demonstrate the necessity for a surface-tension correction in the use of a No. 0 B.S. viscometer. After determining the time of flow (1348 sec.) for water at 25° c. in the usual manner, the water was blown up until the level was about 1 cm. above the upper timing mark and a volume of 0·02 ml. of a solution of " Wettol " (a wetting agent for photographic use supplied by Messrs. Ilford), of density 1·1010 at 25° c., was added at the surface. Flow was then allowed to proceed and the time observed was 1334 sec., approximately 1% less than that for water alone. During the flow the Wettol tended to remain at the air-water interface, and probably none of it reached the capillary during the timed interval. When equilibrium had been reached, the contamination of the water surface was shown by the greatly reduced rise of the meniscus in the capillary above the level of the water in the other arm; the exact rise observed, 2·5 cm., is not significant as the extent of dilution is unknown, but pure water gave a rise of about 7·5 cm. (when the levels in the U-tube were lowered to allow the full rise to occur).

The capillary rise for water in the 2-cm. diameter recipient bulb of a No. 0 viscometer is calculated from Sugden's table as 0·034 cm., and would have been reduced to less than 0·002 cm. if Wettol had been added to this surface also. The surface-tension correction with pure water is thus some 0·7% greater than with a liquid having about 1/3 the surface tension of water.

The correction calculated for water in a bulb of only 3·85 ml. capacity was less than 0·7%; for the 6·5 ml. bulb of a No. 0 viscometer it would be appreciably less, and it cannot have been completely eliminated by the addition of Wettol. Hence some factor, such as the distortion—suggested above—of the meniscus by liquid draining from the wall, must have been operative to produce the observed difference of flow times. An attempt is often made, especially in viscometers, such as that of Bingham, which are operated by external pressure, to eliminate the effect of surface tension by constructing the recipient bulb to be of the same shape and size as the timing bulb; if the falling and rising menisci differ in shape the object cannot be completely achieved.

§ 6. DETERMINATION OR ELIMINATION OF THE SURFACE-TENSION CORRECTION

From the absence of any simple relation between a^2, r and h_σ it will be obvious that the need for a surface-tension correction cannot be avoided by selecting such a diameter for a cylindrical lower bulb as will give a capillary rise equal to the mean capillary rise in the upper bulb: the matching can be correct only for one value of a^2. The same objection applies to Ubbelohde's proposal (1936) to eliminate the correction, in his "suspended level" viscometer, by causing the lower end of the capillary to expand into a hemispherical surface of calculated radius instead of into a horizontal plane.

Two alternatives appear to remain:—

(1) To calibrate viscometers of the usual forms with two or more liquids of known viscosity and known surface tension, using such times of flow as will avoid errors due to kinetic energy, and to state the different constants occurring for different values of a^2; and

(2) To design viscometers so that the maximum surface-tension correction will be negligibly small.

(1) The former procedure presupposes that the viscosities required can be accurately determined by other means. The dynamic viscosity (poises) may be measured in an instrument of the Bingham type, in which external pressures are used that are 30 or more times the variation in hydrostatic pressure in the viscometer during flow. When all the numerous precautions and corrections are taken (including consideration of the "drainage error" which may be marked in this type) accurate results are obtainable, especially if it is necessary only to make them relative to the viscosity of water. The higher pressures used necessitate smaller diameters for the capillary, to minimize kinetic-energy corrections, than in the Ostwald type, and troubles due to suspended matter in the sample and to obstruction of the capillary are exaggerated. For the calibration of an Ostwald-type instrument a further determination, that of density, is required to convert poises to stokes.

(2) If an instrument for comparison of kinematic viscosities can be produced, such that times of flow are proportional to kinematic viscosity to the precision

required, this viscometer, though possibly inconvenient for routine determinations, might be used to determine the relative viscosities required in method (1). Cannon and Fenske (1936), and more recently Cannon (1944), have described "master" viscometers for this purpose, in which the capillary length and mean head are increased to some 45 cm. on the assumption that errors due to surface tension and to kinetic energy will become negligible. Their estimate of the order of the surface-tension correction (0·7 mm. difference in loss of head as between runs with water and oil) suggests that the error in the master instrument should be some 0·15%. But in their fusiform 3 ml. bulb the effect is probably as much as in the 3·85 ml. bulb, for which calculations on the lines indicated above would show an error of more than 0·2%; and the experiment with Wettol points to this being an underestimate. The error would be less if the large diameter of the recipient bulb (3 cm.) were reduced to equal that of the bulb, as in the form described by Cannon for low viscosities. We propose to try the effect of using heads and capillary lengths up to nearly a metre: the correction should then become progressively less, and this will be tested by the variation in the relative times of flow of water and an organic liquid of lower surface tension as the length is increased.

We have found, however, that it is possible to time with good precision the fall of a meniscus in a uniform tube of 1 cm. diameter on which two fine marks have been etched 5 cm. apart. With the same diameter for the other arm of the U-tube the surface-tension correction should be zero, apart from effects of drainage on the shape of the meniscus. The importance of such effects will obviously be much less than in Cannon's "semi-micro" form, which has a bore of 2·5 mm. for the measuring and receiving tubes. The occurrence or absence of sensible correction should again be demonstrable by increasing the length. Since the ratio of initial to final head is greater in this type of instrument than when wide bulbs of similar capacity are provided, it may be desirable, at least in the shorter models, to make use of two recipient tubes in parallel, of the same diameter as the measuring tube. This arrangement, besides reducing the variation in head by a quarter and halving the effect of errors in loading, can be made to avoid the main error due to maladjustment to the vertical of a U-tube viscometer; if the tubes are equidistant from the vertical capillary they are equivalent, as far as concerns the effect of tilt, to a recipient vertically below the upper bulb, as in Cannon and Fenske's "routine" form.

REFERENCES

BARR, 1931. *Viscometry* (Oxford : The University Press), pp. 72, 79, 127.
BARR, 1933. World Petroleum Congress, London.
BÉNARD, 1907. See paragraph 139, Brillouin, *Leçons sur la viscosité* (Paris : Gauthiers-Villars).

BASHFORTH and ADAMS, 1883. *Capillary Action* (Cambridge : The University Press).
CANNON, 1944. *Industr. Engng. Chem. (Anal. Edn.)*, **16**, 708.
CANNON and FENSKE, 1936. *Oil Gas J.*, **34**, 45.
JONES and FORNWALT, 1938. *J. Amer. Chem. Soc.*, **60**, 1683.
JONES and STAUFFER, 1937. *J. Amer. Chem. Soc.*, **59**, 1630.
SUGDEN, 1921. *J. Chem. Soc.*, **119**, 1483.
UBBELOHDE, 1936. *Oel ü. Kohle*, **42**, 949 ; *J. Inst. Pet. Techn.*, **22**, 33.

SOME THERMODYNAMIC RELATIONS OF RIGID HYGROSCOPIC GELS

By F. L. WARBURTON,
Wool Industries Research Association

ABSTRACT. In studying the absorption of liquids by rigid gels, certain thermodynamic relations are required in terms of stress components and generalized co-ordinates. These have been derived using the analytical methods of Gibbs as developed by Guggenheim, and the conditions under which they hold are clearly defined. The analysis includes a new derivation of Kirchhoff's relation for the heat of absorption, in which it is shown to be an exact relation between the isopiestic and isothermal heat of absorption and the isopiestic and isosteric variation of equilibrium humidity with temperature, where the pressures concerned are the external pressures on the gel-liquid phase. The effect of internal stresses existing in a homogeneous gel-liquid phase due to the absorption of the liquid is obtained by assuming that the energy of expansion of the gel is the same for expansion due to absorption as for expansion due to external forces. The effect of external shear stresses on vapour pressure is considered, although it is not required in the case of homogeneous gels, since the analysis is similar to that for the internal energy of expansion, and the effect of such stresses is required when dealing with more complex structures such as wool cells.

§1. INTRODUCTION

BARKAS' work on wood (1945), and in particular on the effect of natural restraints on the wood-water isotherm, has shown the importance of structural rigidity for the vapour pressure isotherms of gels. Wood is an aggregate of cylindrical cells with rigid sheaths which are partly filled with gel substance, and Barkas has been mainly concerned with the effect of the rigid sheath on the water absorption of the gel substance; the restraints on the gel substance are, in this case, external restraints imposed by the sheath. The restraints need not, however, be external restraints, for any energy of swelling of a gel must be reflected in the free energy of the gel-absorbate system, and the isotherm will therefore depend to some extent on the swelling energy. Cassie (1945) has already attempted to estimate the effect of this swelling energy on the water-vapour isotherm of keratin, but his work has been considerably restricted by lack of adequate elastic data for the gel, and of appropriate relationships between the thermodynamic functions and the elastic parameters.

The purpose of this paper is to establish the necessary relations of the thermo-dynamic functions and the elastic parameters, using Gibbs' method. The first requirement is a transformation from the load variables of thermodynamics

to the stress variables of elasticity, and from this the relations of chemical potential and entropy changes to changes of external stress are readily derived. The problem of the relation of chemical potential to the energy of expansion for internal stresses is, however, best approached through the intrinsic energy term, and this and the heat of expansion are derived in terms of the swelling of the gel.

Gibbs' method has the great advantage over the cycle method used by Barkas that it avoids reference to swelling pressures, which are not easily visualized for the internal stresses in the gel itself due to its swelling.

The analysis is only concerned with cases where the stress can be reduced to pressures along the three principal directions and with a fixed amount of gel with varying quantities of absorbed material; the state of the gel is therefore completely specified by the temperature T, the number of moles absorbed, n, and the loads in the three directions $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, or alternatively the stresses X, Y, and Z, and where shear stresses are involved a rectangular block is considered.

§2. RELATION BETWEEN VARIATIONS AT CONSTANT STRESS AND AT CONSTANT LOAD

From the thermodynamic relations expressed in terms of generalized co-ordinates, expressions of the type

$$(\delta B)_{Tn} = (\partial x/\partial A)_{\mathbf{XYZ}}\delta\mathbf{X} + (\partial y/\partial A)_{\mathbf{XYZ}}\delta\mathbf{Y} + (\partial z/\partial A)_{\mathbf{XYZ}}\delta\mathbf{Z} \quad \ldots\ldots(1)$$

can be obtained (e.g., equations (11) and (12) below), where x, y, z are linear dimensions, A is either n or T and B either the chemical potential, μ, of the absorbed substance or the entropy, S.

It is required to express this in terms of $(\partial x/\partial A)_{XYZ}$ and δX etc., the fifth parameter, determining the state of the gel, remaining constant throughout. If the stresses are small compared with the moduli, one can write $\delta\mathbf{X} = yz\,\delta X$ and $(\partial x/\partial A)_{\mathbf{XYZ}} = (\partial x/\partial A)_{XYZ}$ without appreciable error: but in the general case

$$\delta x = \left(\frac{\partial x}{\partial A}\right)_{XYZ}\delta A + \left(\frac{\partial x}{\partial X}\right)_{AYZ}\delta X + \left(\frac{\partial x}{\partial Y}\right)_{AZX}\delta Y + \left(\frac{\partial x}{\partial Z}\right)_{AXY}\delta Z,$$

$$\therefore \left(\frac{\partial x}{\partial A}\right)_{\mathbf{XYZ}} = \left(\frac{\partial x}{\partial A}\right)_{XYZ} + \left(\frac{\partial x}{\partial X}\right)_{AYZ}\left(\frac{\partial X}{\partial A}\right)_{\mathbf{XYZ}} + \left(\frac{\partial x}{\partial Y}\right)_{AZX}\left(\frac{\partial Y}{\partial A}\right)_{\mathbf{XYZ}}$$

$$+ \left(\frac{\partial x}{\partial Z}\right)_{AXY}\left(\frac{\partial Z}{\partial A}\right)_{\mathbf{XYZ}}, \quad \ldots\ldots(2)$$

but $\mathbf{X} = yzX$,

$$\therefore \quad \delta\mathbf{X} = yz\,\delta X + X(y\,\delta z + z\,\delta y).$$

$$\left(\frac{\partial X}{\partial A}\right)_{\mathbf{XYZ}} = -X\left\{\frac{1}{z}\left(\frac{\partial z}{\partial A}\right)_{\mathbf{XYZ}} + \frac{1}{y}\left(\frac{\partial y}{\partial A}\right)_{\mathbf{XYZ}}\right\}, \quad \ldots\ldots(3)$$

also

$$\left(\frac{\partial x}{\partial X}\right)_{AYZ} = -\frac{x}{E_x},$$

$$\left(\frac{\partial x}{\partial Y}\right)_{AZX} = \sigma_{xy}\frac{x}{E_y}, \quad \left.\begin{array}{c}\\\\\\\end{array}\right\} \quad \ldots\ldots(4)$$

$$\left(\frac{\partial x}{\partial Z}\right)_{AXY} = \sigma_{xz}\frac{x}{E_z},$$

where E_x is Young's Modulus in the x direction, σ_{xy} is the ratio of the contraction in the x direction to the extension in the y direction for a stress in the y direction. Since a positive value of X denotes a pressure, whereas in defining E_x tensions are regarded as positive, the signs in equations (4) are interchanged with their usual values.

Substituting from (3) and (4) in (2),

$$\left(\frac{\partial x}{\partial A}\right)_{XYZ} = \left(\frac{\partial x}{\partial A}\right)_{XYZ} + \frac{x}{E_x}X\left\{\frac{1}{z}\left(\frac{\partial z}{\partial A}\right)_{XYZ} + \frac{1}{y}\left(\frac{\partial y}{\partial A}\right)_{XYZ}\right\}$$
$$-\sigma_{xy}\frac{x}{E_y}Y\left\{\frac{1}{x}\left(\frac{\partial x}{\partial A}\right)_{XYZ} + \frac{1}{z}\left(\frac{\partial z}{\partial A}\right)_{XYZ}\right\}$$
$$-\sigma_{xz}\frac{x}{E_z}Z\left\{\frac{1}{x}\left(\frac{\partial x}{\partial A}\right)_{XYZ} + \frac{1}{y}\left(\frac{\partial y}{\partial A}\right)_{XYZ}\right\}, \quad \ldots\ldots(5)$$

or, rearranging and multiplying by $yz\,\delta X$,

$$yz\left(\frac{\partial x}{\partial A}\right)_{XYZ}\delta X = yz\,\delta X\left(1 + \frac{\sigma_{xy}}{E_y}Y + \frac{\sigma_{xz}}{E_z}Z\right)\left(\frac{\partial x}{\partial A}\right)_{XYZ}$$
$$+ zx\,\delta X\left(\frac{\sigma_{xz}}{E_z}Z - \frac{1}{E_x}X\right)\left(\frac{\partial y}{\partial A}\right)_{XYZ} + xy\,\delta X\left(\frac{\sigma_{xy}}{E_y}Y - \frac{1}{E_x}X\right)\left(\frac{\partial z}{\partial A}\right)_{XYZ};$$

similarly

$$zx\left(\frac{\partial y}{\partial A}\right)_{XYZ}\delta Y = yz\,\delta Y\left(\frac{\sigma_{yz}}{E_z}Z - \frac{1}{E_y}Y\right)\left(\frac{\partial x}{\partial A}\right)_{XYZ}$$
$$+ zx\,\delta Y\left(1 + \frac{\sigma_{yx}}{E_x}X + \frac{\sigma_{yz}}{E_z}Z\right)\left(\frac{\partial y}{\partial A}\right)_{XYZ} + xy\,\delta Y(\ldots\ldots)\left(\frac{\partial z}{\partial A}\right)_{XYZ}$$

and

$$xy\left(\frac{\partial z}{\partial A}\right)_{XYZ}\delta Z = yz\,\delta Z\left(\frac{\sigma_{zy}}{E_y}Y - \frac{1}{E_z}Z\right)\left(\frac{\partial x}{\partial A}\right)_{XYZ}$$
$$+ (\ldots\ldots)\left(\frac{\partial y}{\partial A}\right)_{XYZ} + xy\,\delta Z\left(1 + \frac{\sigma_{zx}}{E_x}X + \frac{\sigma_{zy}}{E_y}Y\right)\left(\frac{\partial z}{\partial A}\right)_{XYZ}.$$

Adding the three equations gives

$$yz\left(\frac{\partial x}{\partial A}\right)_{XYZ}\delta X + zx\left(\frac{\partial y}{\partial A}\right)_{XYZ}\delta Y + xy\left(\frac{\partial z}{\partial A}\right)_{XYZ}\delta Z$$
$$= \left(\frac{\partial x}{\partial A}\right)_{XYZ}yz\left\{\delta X - \frac{Y}{E_y}(\delta Y - \sigma_{xy}\delta X - \sigma_{zy}\delta Z) - \frac{Z}{E_z}(\delta Z - \sigma_{xz}\delta X - \sigma_{yz}\delta Y)\right\}$$
$$\left(\frac{\partial y}{\partial A}\right)_{XYZ}zx\left\{\delta Y - \frac{X}{E_x}(\ldots\ldots) - \frac{Z}{E_z}(\ldots\ldots)\right\}$$
$$+ \left(\frac{\partial z}{\partial A}\right)_{XYZ}xy\{\ldots\ldots\}. \quad \ldots\ldots(6)$$

But $\delta X = yz\,\delta X + Xyz\left(\frac{\delta z}{z} + \frac{\delta y}{y}\right)$, where δz and δy are due to arbitary changes in stress at constant A, and

$$\frac{\delta z}{z} = -\frac{1}{E_z}\delta Z + \frac{\sigma_{zy}}{E_y}\delta Y + \frac{\sigma_{zx}}{E_x}\delta X = -(\delta Z - \sigma_{yz}\delta Y - \sigma_{xz}\delta X)/E_z, \quad \ldots\ldots(7)$$

since $\sigma_{zy}/E_y = \sigma_{yz}/E_z$, etc. (Love, 1920, p. 105).

$$\therefore \ \delta X = yz\{\delta X - X(\delta Y - \sigma_{xy}\delta X - \sigma_{zy}\delta Z)/E_y - X(\delta Z - \sigma_{yz}\delta Y - \sigma_{xz}\delta X)/E_z\},$$

and similarly for δY and δZ, so that if $X = Y = Z$

$$\cdots\cdots(8)$$

$$yz(\partial x/\partial A)_{XYZ}\delta X + zx(\partial y/\partial A)_{XYZ}\delta Y + xy(\partial z/\partial A)_{XYZ}\delta Z$$
$$= (\partial x/\partial A)_{XYZ}\delta X + (\partial y/\partial A)_{XYZ}\delta Y + (\partial z/\partial A)_{XYZ}\delta Z. \quad\cdots\cdots(9)$$

Hence when the initial stress is a pure hydrostatic pressure, (1) gives

$$(\delta B)_{Tn} = xyz\left\{\frac{1}{x}\left(\frac{\partial x}{\partial A}\right)_{XYZ}\delta X + \frac{1}{y}\left(\frac{\partial y}{\partial A}\right)_{XYZ}\delta Y + \frac{1}{z}\left(\frac{\partial z}{\partial A}\right)_{XYZ}\delta Z\right\}$$

$$= V\left\{\frac{1}{x}\left(\frac{\partial x}{\partial A}\right)_{XYZ}\delta X + \frac{1}{y}\left(\frac{\partial y}{\partial A}\right)_{XYZ}\delta Y + \frac{1}{z}\left(\frac{\partial z}{\partial A}\right)_{XYZ}\delta Z\right\}. \quad\cdots\cdots(10)$$

When the stress is a pure hydrostatic pressure we can express the variations in B at constant temperature and composition in terms of variation of linear dimensions at constant stress and independent variations in the stress components. When $\delta X = \delta Y = \delta Z = \delta P$ the relation can be obtained more simply from the ordinary thermodynamic relation for fluids (e.g. Guggenheim (1933), equations (151) and (153)) by the relation

$$(1/V)(\partial V/\partial A)_P = (1/x)(\partial x/\partial A)_P + (1/y)(\partial y/\partial A)_P + (1/z)(\partial z/\partial A)_P.$$

When the initial stress involves pure shear stress the error involved is that of neglecting $(X-Y)z\,\delta y + (X-Z)y\,\delta z$ in comparison with δX, which is small if $(X-Y)$ etc. are small compared with the moduli.

§3. DERIVATION OF THE GENERAL RELATIONS IN TERMS OF GENERALIZED CO-ORDINATES

For a rectangular block of gel of the type considered, the general equation is given by Fowler and Guggenheim (1939), p. 60, as $\delta F = -S\delta T - \Sigma X\delta x + \mu\,\delta n$, where F is the Helmholtz free-energy, and if we define G as $F + \Sigma Xx$ by analogy with the definition of Gibbs' function $G*$,

hence
$$\delta G = -S\delta T + \Sigma x\,\delta X + \mu\,\delta n,$$

$$S = -(\partial G/\partial T)_{nXYZ}, \quad x = (\partial G/\partial X)_{nTYZ}. \text{ etc., } \text{ and } \mu = (\partial G/\partial n)_{TXYZ};$$

hence
$$(\partial S/\partial X)_{nTYZ} = -(\partial^2 G/\partial T\partial X) = -(\partial x/\partial T)_{nXYZ}$$

and
$$(\partial \mu/\partial X)_{nTYZ} = (\partial x/\partial n)_{TXYZ}.$$

For variations at constant temperature and composition

$$(\delta S)_{nT} = (\partial S/\partial X)_{nTYZ}\delta X + (\partial S/\partial Y)_{nTZX}\delta Y + (\partial S/\partial Z)_{nTXY}\delta Z,$$

$$= -(\partial x/\partial T)_{nXYZ}\delta X - (\partial y/\partial T)_{nXYZ}\delta Y - (\partial z/\partial T)_{nXYZ}\delta Z$$

and
$$\cdots\cdots(11)$$

$$(\delta \mu)_{nT} = (\partial x/\partial n)_{TXYZ}\delta X + (\partial y/\partial n)_{TXYZ}\delta Y + (\partial z/\partial n)_{TXYZ}\delta Z. \quad\cdots\cdots(12)$$

* G is not identical with G since in the case of a pure hydrostatic pressure $\Sigma Xx = yzXx$ $+zxYy + xyZz = 3PV$.

When the initial stress is a pure hydrostatic pressure these can be transformed by the relations obtained in the previous section:

$$(\delta S)_{nT} = -V\left\{\frac{1}{x}\left(\frac{\partial x}{\partial T}\right)_{nXYZ}\delta X + \frac{1}{y}\left(\frac{\partial y}{\partial T}\right)_{nXYZ}\delta Y + \frac{1}{z}\left(\frac{\partial z}{\partial T}\right)_{nXYZ}\delta Z\right\}, \quad\ldots\ldots(11.1)$$

$$(\delta\mu)_{nT} = V\left\{\frac{1}{x}\left(\frac{\partial x}{\partial n}\right)_{TXYZ}\delta X + \frac{1}{y}\left(\frac{\partial y}{\partial n}\right)_{TXYZ}\delta Y + \frac{1}{z}\left(\frac{\partial z}{\partial n}\right)_{TXYZ}\delta Z\right\}, \quad\ldots\ldots(12.1)$$

but $(1/x)(\partial x/\partial T)_{nXYZ}$ is the coefficient of thermal expansion in the x direction, α_x and $(1/x)(\partial x/\partial n)_{TXYZ}$ is the isothermal swelling per mole in the x direction at constant pressure $= s_x$; hence

$$(\delta S)_{nT} = -V(\alpha_x\delta X + \alpha_y\delta Y + \alpha_z\delta Z) \qquad\ldots\ldots(11.2)$$

and

$$(\delta\mu)_{nT} = V(s_x\delta X + s_y\delta Y + s_z\delta Z). \qquad\ldots\ldots(12.2)$$

Equation (12.2) can be used to calculate the effect of shear stress on vapour pressure, and the variation of vapour pressure of a hygroscopic gel initially subject to a pure hydrostatic pressure (or when shear stresses are small compared with the rigidity) with any variation in its state. Equation (11.2) can be used to calculate the contribution of the energy of expansion to the heat of absorption of the moisture by the gel. These will be considered separately.

§4. EFFECT OF CHANGES IN STATE OF A GEL INITIALLY SUBJECT TO A PURE HYDROSTATIC PRESSURE ON THE EQUILIBRIUM VAPOUR PRESSURE

Equation (12.2) gives the variation in μ at constant temperature and composition with changes in the stress components; the net variation when temperature and composition may also change is therefore given by

$$\delta\mu = (\partial\mu/\partial n)_{PT}\delta n + (\partial\mu/\partial T)_{Pn}\delta T + V(s_x\delta X + s_y\delta Y + s_z\delta Z), \quad\ldots\ldots(13)$$

since constant X, Y and Z when $X=Y=Z$ is identical with constant pressure. If μ_V is the chemical potential of the vapour in equilibrium with the absorbed component in the gel, $\mu_V = \mu$ at all times.

The only parameters on which μ_V depends are the vapour pressure p and the temperature T. Equilibrium can only be maintained if T is the same for both phases.

$$\therefore\ \delta\mu = \delta\mu_V = (\partial\mu_V/\partial p)_T\delta p + (\partial\mu_V/\partial T)_p\delta T. \qquad\ldots\ldots(14)$$

Combining (13) and (14),

$$(\partial\mu/\partial n)_{PT}\delta n + (\partial\mu/\partial T)_{Pn}\delta T + V(s_x\delta X + s_y\delta Y + s_z\delta Z) = (\partial\mu_V/\partial p)_T\delta p + (\partial\mu_V/\partial T)_p\delta T.$$

We have (Guggenheim, 1933, p. 35)

$$(\partial\mu/\partial T)_{Pn} = -(\partial S/\partial n)_{PT}, \qquad (\partial\mu_V/\partial p)_T = v, \qquad (\partial\mu_V/\partial T)_p = -S_V,$$

where $v = $ molar volume of the vapour.

$$\therefore\ (\partial\mu/\partial n)_{PT}\delta n - (\partial S/\partial n)_{PT}\delta T + V(s_x\delta X + s_y\delta Y + s_z\delta Z) = v\delta p - S_V\delta T.$$

$$\ldots\ldots(15)$$

Considering first isothermal changes at constant stress:

$$(\partial\mu/\partial n)_{PT}\delta n = v\,\delta p$$

$$\therefore\ (\partial\mu/\partial n)_{PT} = v(\partial p/\partial n)_{PT}.$$

The general equation for isothermal changes is

$$V(s_x\delta X + s_y\delta Y + s_z\delta Z) = v\,\delta p - v(\partial p/\partial n)_{XYZT}\delta n. \qquad \ldots\ldots(16)$$

This is identical with equation (13) in the paper by Barkas (1945) on swelling stresses in rigid gels, but the derivation given here shows that it is only rigid if the initial stress is a pure hydrostatic pressure. This can also be written

$$V\left\{(s_x+s_y+s_z)\left(\frac{\delta X+\delta Y+\delta Z}{3}\right) + (s_x-s_y)\left(\frac{\delta X-\delta Y}{3}\right)\right.$$
$$\left. + (s_y-s_z)\left(\frac{\delta Y-\delta Z}{3}\right) + (s_z-s_x)\left(\frac{\delta Z-\delta X}{3}\right)\right\} = s\,\delta p - v\left(\frac{\partial p}{\partial n}\right)_{XYZT}\delta n,$$

thus separating the effect of changes in hydrostatic pressure and shear stresses.

Equation (15) can also be used to determine the variation of vapour pressure with temperature keeping the composition constant. This can be done even if the absorbed liquid is in more than one form, as equilibrium must be maintained between these forms, and, therefore, according to the phase rule, in discussing equilibrium with external conditions the absorbed liquid can be treated as a single component. In this case the only variation on the L.H.S. of equation (15) is in T, whereas for the vapour phase the pressure must be allowed to alter to maintain equilibrium.

$$\therefore \quad -\left(\frac{\partial S}{\partial n}\right)_{PT}\delta T = v\,\delta p - S_V\delta T,$$

$$\therefore \quad S_V-\left(\frac{\partial S}{\partial n}\right)_{PT} = v\left(\frac{\partial p}{\partial T}\right)_{nP} = \frac{RT}{p}\left(\frac{\partial p}{\partial T}\right)_{nP}$$

(if the vapour behaves as a perfect gas);

$$\therefore \quad T\left\{S_V-\left(\frac{\partial S}{\partial n}\right)_{PT}\right\} = RT^2\left(\frac{\partial}{\partial T}\ln p\right)_{nP},$$

but

$$T\left(\frac{\partial S}{\partial n}\right)_{PT} = \left(\frac{\partial H}{\partial n}\right)_{PT} - \mu$$

(Guggenheim, pp. 37 and 38), where H is the heat content of the gel ;

$$\therefore \quad H_V-\left(\frac{\partial H}{\partial n}\right)_{PT} + (\mu-\mu_V) = RT^2\left(\frac{\partial}{\partial T}\ln p\right)_{nP},$$

but $\mu = \mu_V$,

$$\therefore \quad H_V-\left(\frac{\partial H}{\partial n}\right)_{PT} = RT^2\left(\frac{\partial}{\partial T}\ln p\right)_{nP}. \qquad \ldots\ldots(17)$$

This may be combined with the relation obtained by Guggenheim (1933, p. 64), using a similar analysis, for the equilibrium vapour pressure of a pure liquid maintained at constant hydrostatic pressure, which may be written

$$H_V-H_L = RT^2\left(\frac{\partial}{\partial T}\ln p_0\right)_{nP}, \qquad \ldots\ldots(18)$$

where H_L is the molar total heat of the liquid and p_0 the pressure of the vapour

in equilibrium with the pure liquid. Whence—since for a perfect gas H_V is independent of the pressure at constant temperature—

$$H_L - \left(\frac{\partial H}{\partial n}\right)_{PT} = RT^2 \left\{ \frac{\partial}{\partial T} \ln\left(\frac{p}{p_0}\right) \right\}_{nP}.$$

The heat absorbed when 1 mole of liquid is absorbed by a large quantity of gel at constant temperature and pressure is given by the increase in total heat

$$\left(\frac{\partial}{\partial n} \Delta H\right)_{PT},$$

where

$$(\partial \Delta H/\partial n)_{PT} = (\partial H/\partial n)_{PT} - H_L.$$

$$\therefore \quad \left(\frac{\partial}{\partial n} \Delta H\right)_{PT} = -RT^2 \left(\frac{\partial}{\partial T} \ln \frac{p}{p_0}\right)_{nP} = -RT^2 \left(\frac{\partial}{\partial T} \ln h\right)_{nP}, \quad \ldots\ldots(19)$$

where h is the relative humidity of the vapour in equilibrium with the liquid absorbed by the gel.

The above equation is identical with Kirchhoff's relation for the heat of dilution of a dilute solution, but the above derivation shows that it is perfectly general and makes clear precisely what quantity is given by $RT^2(\partial/\partial T \ln h)_{nP}$. Previous derivations of the relation by Shorter (1924) and Hedges (1926) have either directly applied the relation for dilute solutions on general thermodynamic reasoning or have been based on the assumption that the volume swelling per mole was identical with the molar volume of the pure liquid, which does not hold for keratin at low regains. It should be noted that the heat of absorption Q as defined by Hedges and Shorter $= -(\partial \Delta H/\partial n)/M$, where M is the molecular weight.

Equation (17) has also been derived by earlier workers, notably Kruyt and Modderman (1930) and Brunauer (1943), who applied it to the heat of absorption of gases by charcoal, etc., and who give it as $RT^2(\partial/\partial T \ln p)_n = q$, where q is variously defined, whereas in equation (17) the heat is accurately defined in terms of generally understood thermodynamic quantities. The relation of this to the observed heat of absorption in any experimental arrangement is a matter for thermodynamic analysis of the exact procedure employed, and will differ for different experimental arrangements.

In the case of direct absorption by a gel from the pure liquid, the usual procedure is to immerse a given quantity of the gel containing n moles of the absorbed liquid in a calorimeter containing the pure liquid and observe the heat evolved. Since in this case the pressure on the liquid and the gel remain constant throughout and the rise in temperature is negligibly small (although the quantity actually measured), the heat evolved is clearly,

$$-\int_n^{\text{sat}} \left(\frac{\partial \Delta H}{\partial n}\right)_{PT} dn,$$

where the integral covers the range from the initial content of absorbate to saturation. For an amount of gel containing $100M$ grams of the dry substance, where M is the molecular weight of the absorbed liquid, n is the regain C

(amount of absorbate per 100 g. of dry substance), and hence the amount of heat evolved per gram of the dry substance is

$$-(1/100M)\int_0^{sat} \left(\frac{\partial \Delta H}{\partial n}\right)_{PT} dC$$

$$= -(1/1800)\int_0^{sat} \left(\frac{\partial \Delta H}{\partial n}\right)_{PT} dC \text{ for water.} \qquad \ldots\ldots(20)$$

In the case of direct absorption of the pure liquid by the gel, the heat evolved is thus directly related to the quantity given by Kirchhoff's relation, which has been shown to be an exact relation between the heat absorbed when 1 mole of absorbed substance is absorbed by a large quantity of gel material at constant temperature and pressure on the one hand, and the isosteric variation of equilibrium humidity at constant hydrostatic pressure on the gel, on the other hand.

§5. EFFECT OF GENERAL EXPANSION OF GEL ON THE INTRINSIC ENERGY AND THE CONTRIBUTION OF THE ENERGY OF EXPANSION TO THE TOTAL HEAT OF ABSORPTION

Equation (11.2) expresses the effect on the entropy of varying the stress components on a rigid gel initially under a pure hydrostatic pressure. This can be used to calculate the change in intrinsic energy δU, when the gel is expanded by the application of external forces by the same amount as when an additional mole of liquid is absorbed at constant temperature and pressure. Experiment does not directly give the changes in stress required to do this, and equation (11.2) must be transformed by expressing the changes in stress in terms of changes in the geometric parameters x, y, and z. This can be done by means of the ordinary elastic equations (for rhombic symmetry):

$$\left.\begin{aligned}
-\frac{\delta x}{x} &= \frac{\delta X}{E_x} - \sigma_{xy}\frac{\delta Y}{E_y} - \sigma_{xz}\frac{\delta Z}{E_z}, \\
-\frac{\delta y}{y} &= -\sigma_{yx}\frac{\delta X}{E_x} + \frac{\delta Y}{E_y} - \sigma_{yz}\frac{\delta Z}{E_z}, \\
-\frac{\delta z}{z} &= -\sigma_{zx}\frac{\delta X}{E_x} - \sigma_{yz}\frac{\delta Y}{E_y} + \frac{\delta Z}{E_z}.
\end{aligned}\right\} \qquad \ldots\ldots(21)$$

The negative sign appears on the L.H.S., as the conventional definition of E assumes that tensions are positives, whereas in the thermodynamic convention tensions are regarded as negative pressures.

These three equations can be transformed to another set expressing the stresses in terms of the resulting strains. Writing A_x for $\delta X/E_x$, this becomes

$$\left.\begin{aligned}
-DA_x &= (1 - \sigma_{zy}\sigma_{yz})\,\delta x/x + (\sigma_{xy} + \sigma_{xz}\sigma_{zy})\,\delta y/y + (\sigma_{xz} + \sigma_{xy}\sigma_{yz})\,\delta z/z, \\
-DA_y &= (1 - \sigma_{xz}\sigma_{zx})\,\delta y/y + (\sigma_{yz} + \sigma_{yx}\sigma_{xz})\,\delta z/z + (\sigma_{yx} + \sigma_{yz}\sigma_{zx})\,\delta x/x, \\
-DA_z &= (1 - \sigma_{xy}\sigma_{yx})\,\delta z/z + (\sigma_{zx} + \sigma_{zy}\sigma_{yx})\,\delta x/x + (\sigma_{zy} + \sigma_{zx}\sigma_{xy})\,\delta y/y,
\end{aligned}\right\} \qquad \ldots\ldots(22)$$

where D is a determinant of which the value is

$$D = 1 - \sigma_{xy}\sigma_{yx} - \sigma_{xz}\sigma_{zx} - \sigma_{yz}\sigma_{zy} - \sigma_{xy}\sigma_{yz}\sigma_{zx} - \sigma_{yx}\sigma_{zy}\sigma_{xz}.$$

A similar transformation for colour equations is given by J. Guild (1924). These values can be substituted in equation (12.2), giving

$$(\delta S)_{nT} = (V/D)[E_x\alpha_x\{(1-\sigma_{zy}\sigma_{yz})\,\delta x/x + (\sigma_{xy}+\sigma_{xz}\sigma_{zy})\,\delta y/y$$
$$+ (\sigma_{xz}+\sigma_{xy}\sigma_{yz})\,\delta z/z\} + E_y\alpha_y\{(1-\sigma_{xz}\sigma_{zx})\delta y/y + \ldots\} + E_z\alpha_z\{\ldots\ldots\}].$$
$$\ldots\ldots(23)$$

If U is the intrinsic energy, $\delta U = T\delta S - \Sigma X\,\delta x$ at constant composition; or, when the initial stress is pure hydrostatic pressure,

$$\delta U = T\delta S - P\delta V;$$

$$\therefore\quad (\delta U)_{nT} = (TV/D)[E_x\alpha_x\{\ldots\ldots\} + E_y\alpha_y\{\ldots\ldots\} + E_z\alpha_z\{\ldots\ldots\}] - P\delta V.$$

If we assume that the internal energy of differential expansion is the same whether the expansion is due to external forces or absorption of liquid,[*] the contribution of the expansion to the total heat or, more strictly, the total heat of absorption, when there is no heat of reaction between the absorbed liquid and the gel when absorption takes place at constant temperature and pressure, is given by

$$(\delta H)_{\text{expansion}} = (TV/D)[E_x\alpha_x\{\ldots\ldots\} + E_y\alpha_y\{\ldots\ldots\} + E_z\alpha_z\{\ldots\ldots\}] - P\delta V + P\delta V,$$

since at constant pressure $\delta(PV) = P\delta V$. Expressing this per mole,

$$(\partial H/\partial n)_{\text{expansion}} = (TV/D)[E_x\alpha_x\{(1-\sigma_{zy}\sigma_{yz})(1/x)(\partial x/\partial n)_{PT} + \ldots\ldots\} + \ldots\ldots]$$
$$= (TV/D)[E_x\alpha_x\{(1-\sigma_{zy}\sigma_{yz})s_x + (\sigma_{xy}+\sigma_{xz}\sigma_{zy})s_y + \ldots\ldots\} + \ldots\ldots],$$
$$\ldots\ldots(24)$$

writing s_x for $(\partial x/\partial n)/x$, etc., in each of the terms in the equation.

If there is symmetry about the z axis,

$$\sigma_{xy} = \sigma_{yx} = \sigma_{xx},$$
$$\sigma_{zy} = \sigma_{zx},$$
$$\sigma_{yz} = \sigma_{xz},$$

and $\quad s_y = s_x, \quad \alpha_x = \alpha_y, \quad$ and $E_x = E_y.$

The expression for D then reduces to

$$D = (1+\sigma_{xx})(1-\sigma_{xx}-2\sigma_{xz}\sigma_{zx})$$

and

$$(\partial H/\partial n)_{\text{expansion}} = (TV/D)[E_x\alpha_x\{(1-\sigma_{xz}\sigma_{zx})s_x + (\sigma_{xx}+\sigma_{xz}\sigma_{zx})s_x$$
$$+ (\sigma_{xz}+\sigma_{xx}\sigma_{xz})s_z\} + E_x\alpha_x\{\ldots\ldots\} + E_z\alpha_z\{(1-\sigma_{xx}^2)s_z + 2(\sigma_{zx}+\sigma_{zx}\sigma_{xx})s_x\}]$$
$$= [TV/(1+\sigma_{xx})(1-\sigma_{xx}-2\sigma_{xz}\sigma_{zx})][2E_x\alpha_x\{(1+\sigma_{xx})s_x + \sigma_{xz}(1+\sigma_{xx})s_z\}$$
$$+ E_z\alpha_z\{(1+\sigma_{xx})(1-\sigma_{xx})s_z + 2\sigma_{zx}(1+\sigma_{xx})s_x\}]$$
$$= [TV/(1-\sigma_{xx}-2\sigma_{xz}\sigma_{zx})][2E_x\alpha_x(s_x+\sigma_{xz}s_z) + E_z\alpha_z\{2\sigma_{zx}s_x + (1-\sigma_{xx})s_z\}],$$
$$\ldots\ldots(25)$$

but $\qquad\qquad E_x\sigma_{xz} = E_z\sigma_{zx},$

[*] This assumption is the simplest one, and would appear to be justified by the fact that it gives heats of the correct order when this is the principal source of change in intrinsic energy on absorption of water.

therefore an alternative form is

$$(\partial H/\partial n)_{\text{expansion}} = \{TV/(1 - \sigma_{xx} - 2\sigma_{xz}\sigma_{zx})\}[2E_x\{\alpha_x s_x$$
$$+ \sigma_{xz}(\alpha_x s_z + \alpha_z s_x)\} + E_z\alpha_z s_z(1 - \sigma_{xx})]. \quad \ldots\ldots(26)$$

For most hygroscopic gels s_z is only about 3% of s_x, and it is probable that the same is true of α_z and α_x (as it certainly is for wood). In this case, the last term in the square brackets can be omitted and we have

$$(\partial H/\partial n)_{\text{expansion}} = 2TVE_x\{\alpha_x s_x + \sigma_{xz}(\alpha_x s_z + \alpha_z s_x)\}/(1 - \sigma_{xx} - 2\sigma_{zx}\sigma_{xz}).$$
$$\ldots\ldots(27)$$

It should be emphasized that the above analysis assumes that the gel is homogeneous in the limited sense that external stresses can produce the same distortions at every point as absorption of the liquid.

§6. APPLICATION TO THE STUDY OF THE ABSORPTION OF WATER BY TEXTILE FIBRES

Urquhart and Williams (1924) determined the variation in the equilibrium vapour pressure of cotton with temperature and regain and give the results in the form of a series of iso-humidity curves. From these it is seen that at high relative humidities the equilibrium regain passes through a minimum as the temperature increases, the temperature corresponding to this minimum increasing as the humidity decreases. At these minima the relative humidity in equilibrium with the corresponding regain is a maximum, i.e. $(\partial/\partial T \ln h)_{nP}$ is positive for lower temperatures and negative for higher temperatures. Hence, from equation (19), $(\partial\Delta H/\partial n)_{PT}$ for the particular high regain concerned is negative for lower temperatures and positive for high temperatures.

A somewhat similar result was obtained by Speakman, Stott and Chang (1933) for wool. They found that the saturation regain and that for 97·5% R.H. showed a minimum at temperatures between 40 and 50° c. and also, as might be expected, that the swelling in water was a minimum at about the same temperature. The change in sign of the value of $(\partial\Delta H/\partial n)_{PT}$ with increase in temperature is most easily explained by the assumption that $(\partial\Delta H/\partial n)_{PT}$ consists of two components of opposite sign, and that either the value of each changes with temperature or the contribution of each to the total value of $(\partial\Delta H/\partial n)_{PT}$ changes. Such an assumption is made by Cassie (1945), who attributes the value of $(\partial\Delta H/\partial n)_{PT}$ as observed by Hedges to two components—a negative component changing rapidly with regain, due to the heat of reaction between the gel and the water, and a positive component due to the heat required to expand the gel. Combining this assumption with the theory of Peirce (1929) and others, that the absorbed water consists of two portions, one of which has no heat of reaction, it is seen that $(\partial\Delta H/\partial n)_{PT}$ for this portion is positive, while for the remainder it is negative. If equation (19) is applied separately to these two portions it is seen that the relative humidity in equilibrium with one portion decreases with increase in temperature, and for the other portion increases, if the proportions of the two components remain the same. Equilibrium can in fact only be maintained by taking into account the first term on the L.H.S. of equation (15), which will have

two components, and adding the relation that the sum of the changes in δn for the two portions is zero. When this is done, equation (19) is found to hold for the water as a whole, as deduced from the phase rule, but the proportion of the component with a positive value of $(\partial \Delta H / \partial n)_{PT}$ will increase with rise in temperature, thus tending to make $(\partial \Delta H / \partial n)_{PT}$, for the water as a whole, become less negative as the temperature rises, in accordance with the experimental results quoted.

At high regains and temperatures, when practically all the water is absorbed without heat of reaction, $(\partial \Delta H / \partial n)_{PT}$ should approach the heat of expansion of the gel as given by equation (27). However, the comparison requires a knowledge of the transverse value of Young's Modulus, E_x, as well as the three principal values of Poisson's ratio for a material showing radial symmetry. Unfortunately it is impossible to obtain these constants for textile fibres, and an accurate comparison can only be made by the use of a material like horn keratin, which is obtainable in bulk, and shows similar absorption properties to wool keratin. Until full experimental data for horn keratin are available, the best comparison that can be made is between $(\partial \Delta H / \partial n)_{PT}$, calculated from Urquhart and Williams' isostere for cotton at 19·9% regain, which corresponds to about 98 % R.H., and for which full experimental values are given, and $(\partial H / \partial n)_{\text{expansion}}$ calculated on certain assumptions for the value of the elastic constants, etc., of cotton. The assumptions are:

(1) E_x for cotton is about 10^4 bars, which is approximately the value for horn keratin and the gel material of wood at high regains (W.I.R.A. and Barkas, 1945, p. 55).

(2) α_x is that for beech across the grain, $= 6 \times 10^{-5}$ (Kaye and Laby).

(3) $2V\{s_x + \sigma_{xz}(s_z + s_x \alpha_z / \alpha_x)\} = $ molar volume of water $= 18$ c.c.

(4) $\sigma_{xz} = \sigma_{xz} = \sigma_{zx} = \frac{1}{4}$ (value of σ_{xz} for keratin (W.I.R.A.)).

Substituting in equation (27) for $T = 360°$ K. $= 87°$ C.,

$$(\partial H / \partial n)_{\text{expansion}} = \frac{360 \times 10^{10} \times 18 \times 6 \times 10^{-5}}{1 - \frac{1}{4} - \frac{1}{8}} \times 2 \cdot 4 \times 10^{-8} \text{ cals/mole}$$

$$= 160 \text{ cal/mole.}$$

Urquhart and Williams' data give the equilibrium vapour pressure of cotton at 19·9% regain when it is absorbing water at three temperatures, 70, 80, and 90° C. These can be converted into equilibrium humidities with the aid of saturation vapour pressures at these temperatures as given by the Smithsonian Meteorological Tables and the logarithm of h plotted against the reciprocal of the temperature. $(\partial \Delta H / \partial n)$ at any temperature is then given by the slope of the curve. The value found at 90° C. is 64 cal/mole, which is of the same order of magnitude as the heat of expansion obtained from elastic constants.

Urquhart and Williams also give the absorption and desorption isosteres at 12% regain in the form of curves of the logarithm of the vapour pressure against the reciprocal of the temperature, obtaining a greater (negative) slope for desorption than absorption, showing that $(\partial H / \partial n)_{PT}$ is less positive for desorption than absorption.

This is consistent with the assumption of Barkas (1942) and Cassie (1945) that the hysteresis of the absorption isotherm is a reflection of the elastic hysteresis of the gel, as on this assumption the heat of expansion will be less for desorption than absorption.

The absorption isotherm gives the change in the thermodynamic potential or Gibbs' function when one mole of the absorbate is transferred from the liquid in bulk to a large quantity of the gel at constant temperature and pressure. The general expression for the change in this function is given as

$$\Delta G = \Delta H - T\Delta S;$$

differentiating with respect to n gives the change in chemical potential under the conditions mentioned above, i.e.,

$$\Delta \mu = (\partial \Delta G / \partial n)_{PT} = (\partial \Delta H / \partial n)_{PT} - T(\partial \Delta S / \partial n)_{PT}.$$

The study of the isotherm is thus resolved into a study of the two components of the change in Gibbs' function. The analysis of the entropy term involves statistical considerations and is outside the scope of this paper. The energy term can be obtained experimentally either from the heat of wetting or from the isostere, and can therefore be analysed independently of the entropy term. The change in total heat ΔH can be written

$$\Delta H = \Delta U + \Delta PV$$

and the change in intrinsic energy can be split into two portions: ΔU_1 the change in intrinsic energy due to the reaction between the liquid and the gel, and ΔU_2 the change due to the expansion of the gel. The ΔPV term involves the change in the value of PV for the gel, and also for the liquid, so that, differentiating,

$$\left(\frac{\partial \Delta H}{\partial n}\right)_{PT} = \left(\frac{\partial \Delta U_1}{\partial n}\right)_{PT} + \left(\frac{\partial \Delta U_2}{\partial n}\right)_{PT} + P\left(\frac{\partial V}{\partial n}\right)_{PT} - PV_L, \quad \ldots\ldots(28)$$

where $(\partial V / \partial n)_{PT}$ refers to the gel and V_L is the molar volume of the liquid in bulk. The second term is that obtained from the elastic properties of the gel, and hence the second and third terms are equivalent to $(\partial H / \partial n)_{\text{expansion}}$. If the absorbed liquid takes two forms with different energies of reaction, the differentiation takes place under conditions of equilibrium between the two forms. It is generally assumed that the liquid in one of these forms is localized, and in the other form relatively mobile. If A is the number of moles of liquid in the mobile form and B the number in the localized (bound) form,

$$\left(\frac{\partial \Delta U_1}{\partial n}\right)_{PT} = \left(\frac{\partial \Delta U_1}{\partial A}\right)_{BPT}\left(\frac{\partial A}{\partial n}\right)_{PT} + \left(\frac{\partial \Delta U_1}{\partial B}\right)_{APT}\left(\frac{\partial B}{\partial n}\right)_{PT}, \quad \ldots\ldots(29)$$

$$\therefore \left(\frac{\partial \Delta H}{\partial n}\right)_{PT} = \left(\frac{\partial \Delta U_1}{\partial A}\right)_{BPT}\left(\frac{\partial A}{\partial n}\right)_{PT} + \left(\frac{\partial \Delta U_1}{\partial B}\right)_{APT}\left(\frac{\partial B}{\partial n}\right)_{PT} + \left(\frac{\partial H}{\partial n}\right)_{\text{expansion}} - PV_L.$$
$$\ldots\ldots(30)$$

At low contents of the absorbate it is generally assumed that all the liquid is absorbed in the bound form, i.e. when $n = 0$,

$$\left(\frac{\partial A}{\partial n}\right)_{PT} = 0 \quad \text{and} \quad \left(\frac{\partial B}{\partial n}\right)_{PT} = 1;$$

$$\therefore \quad \left(\frac{\partial \Delta H}{\partial n}\right)_{PT, n=0} = \left(\frac{\partial \Delta U_1}{\partial B}\right)_{APT} + \left(\frac{\partial H}{\partial n}\right)_{\text{expansion}} - PV_L. \quad \ldots\ldots(31)$$

Hence the value of $(\partial \Delta U_1/\partial B)$ can be determined at low contents. If it is further assumed, as has been done by Cassie (1945), following Pierce (1929), that $(\partial \Delta U_1/\partial A)$ is zero and that $(\partial \Delta U_1/\partial B)$ is a constant, substitution of the value of the latter in equation (30) enables $\partial B/\partial n$ to be found for all values of n. The values of B so obtained can be used in analysing the entropy term on statistical and other theories.

ACKNOWLEDGMENTS

REFERENCES

BARKAS, W. W., 1942. *Trans. Faraday Soc.*, 38, 207.
BARKAS, W. W., 1945. *Swelling Stresses in Gels*, F.P.R. Special Report, No. 6 (H.M.S.O.)
BRUNAUER, S., 1943. *The Adsorption of Gases and Vapours* (Oxford: The University Press), 1, 222.
CASSIE, A. B. D., 1945. *Trans. Faraday Soc.*, 41, 458–64.
FOWLER, R. H. and GUGGENHEIM, E. A., 1939. *Statistical Thermodynamics* (Cambridge: The University Press).
GUGGENHEIM, E. A., 1933. *Modern Thermodynamics* (London: Methuen & Co.).
GUILD, J., 1924–5. *Trans. Opt. Soc.*, 26, 99 and 100.
HEDGES, J. J., 1926. *Trans. Faraday Soc.*, 22, 178.
KAYE and LABY, 1936. *Physical and Chemical Constants*, 8th Edition (London: Longmans & Co.).
KRUYT, H. R. and MODDERMAN, J. G., 1930. *Chem. Rev.* 7, 278–82.
LOVE, A. E. H., 1920. *Mathematical Theory of Elasticity*, 3rd Edition (Cambridge: The University Press).
PEIRCE, F. T., 1929. *J. Text. Inst.*, 20, T. 133.
SHORTER, S. A., 1924. *J. Text. Inst.*, 15, T. 328.
SPEAKMAN, J. B., STOTT, E. and CHANG, H., 1933. *J. Text. Inst.*, 24, T. 284.
URQUHART, A. R. and WILLIAMS, A. M., 1924. *J. Text. Inst.*, 15, T. 559–72.
W.I.R.A. Unpublished work.

OPTICAL PROBLEMS OF THE ROTATING-PRISM CINEMATOGRAPH PROJECTOR

By J. KUDAR

ABSTRACT. The most characteristic optical aberration in the rotating-prism non-intermittent cinematograph projector, the non-linearity of the image shift, can be eliminated by a suitable periodic variation of the direction of the illuminating beam. In addition to a description of this unusual method of optical correction, other aberrations of the image formation by the rotating prism are discussed, such as the prismatic astigmatism and the tangential and sagittal prismatic coma.

§1. INTRODUCTION

IN cinematograph projectors with non-intermittent film movement, an optical device is used for producing a stationary projected image from the uniformly moving film. This may be an arrangement of rotating or oscillating mirrors, or a rotating polygonal prism. A ray suffers parallel displacement by refraction in a plane parallel plate, represented by opposite sides of the rotating prism (cf. figure 1); a polygonal prism of suitable dimensions and speed of rotation can, therefore, produce a stationary image. However, the parallel shift is not a linear function of the angle of incidence, and this gives rise to considerable optical aberrations. A theory of these was worked out by H. Dennis Taylor (1937) and applied especially for high-speed cinematographic cameras with rotating prism. The use of such cameras in practice is made possible by the circumstance that the light is cut off by a rotating shutter between two consecutive images in those positions of the prism where the angles of incidence, and thus the aberrations, would be too large. This limitation of the angle of incidence is impracticable in the case of projectors, because it would involve a frequency of interruption equal to the number of images per second. With the usual number of 24 images/sec., this would cause troublesome flicker*; a dark period in the middle of the projection of an image, applied for doubling the flicker frequency, would cut out most of the time available for the projection of the image if its length were made comparable with that dark period between two images which would result from the exclusion of large angles. In addition, it would result in the loss of the optically most perfect interval. Increasing the number of sides of the polygon would reduce the greatest angle of incidence, but the natural limitation to the circumference of the rotating prism prevents sufficiently good correction by this means. It seems more promising, therefore, to investigate optical possibilities for correcting or compensating the non-linearity of the image shift.

Detailed consideration shows that, for practical reasons, it would be very difficult to attempt correction by means of rotating or oscillating optical elements between the film and the projection objective or in front of the objective. Fortun-

---

* With modern intermittent projectors, the number of interruptions is two or three times the number of images.

ately, however, the non-linearity of the image shift can be eliminated by a very unusual method of correction, without introducing any constructional parts between the film and the objective or in front of the objective. This can be understood by means of figures 1 and 2. Figure 1 shows small pencils of rays coming from the film F parallel to the optic axis A–A of the projection objective O and deviated by refraction in the prism (of which p and p′ are opposite parallel sides).

§2. DETAILED THEORY

Let D be the distance between the sides p and p′, n the refractive index of the prism, x the angle of incidence which is equal to the angle between the objective axis and the prism face normal and y the angle of refraction. The parallel shift h of the rays due to the refraction is

$$h = \frac{D\sin(x-y)}{\cos y} = D\sin x\left(1 - \frac{\cos x}{\sqrt{(n^2 - \sin^2 x)}}\right) = D\frac{n-1}{n}\left[x + \left(\frac{n+1}{2n^2} - \frac{1}{6}\right)x^3 + \dots\right],$$

$$\dots\dots(1)$$

where the terms in x^5 and higher powers of x can be neglected for practical purposes. Equation (1) shows that the non-linear aberration of the parallel shift, represented by the term with x^3, is positive. A small pencil from any point M of the film produces an astigmatic image M′–M″ which is nearer to the optic axis by the distance (1) than the original point M.

Let now figure 2 represent another case in which narrow parallel pencils of the illuminating rays remain perpendicular, not to the film, but to the face p of the rotating prism. Let again x be the angle between the axis of the projection and the prism face normal. In this case there is no parallel shift, but, seen from the objective O, a point M of the film appears as an anastigmatic image M′ produced by the refraction at the prism. For pencils of small aperture the distance between M and M′ is

$$MM' = D\frac{n-1}{n}. \qquad\dots\dots(2)$$

Relatively to the point M, the image M′ is nearer to the objective axis by the distance

$$h' = MM' \cdot \sin x = D\frac{n-1}{n}\left(x - \frac{x^3}{6} + \dots\right). \qquad\dots\dots(3)$$

Equation (3) represents the compensation of the movement of the film in the case shown by figure 2, while equation (1) represents the same for the case of figure 1. Both cases involve non-linear aberrations, as shown by the presence of the terms with x^3; the essential difference, however, is that the sign of the aberration is different.

Figures 1 and 2 represent extreme cases in which the movement of the film is over-compensated or under-compensated. By a suitable choice of conditions, therefore, an intermediate case can be obtained in which the aberration term in x^3 disappears. The condition for this is provided by the following rule: the ratio of the angle of incidence of the principal ray illuminating a point of the film to the angle between the prism face normal and the objective axis must be constant during the rotation of the prism. This ratio is 1 in the case of figure 1, and 0 in the

case of figure 2; the non-linear aberration disappears at an intermediate value
which depends on the refractive index of the prism.

This optical condition for the elimination of the non-linear aberration of the
image shift can be fulfilled, for example, by introducing between the film and the
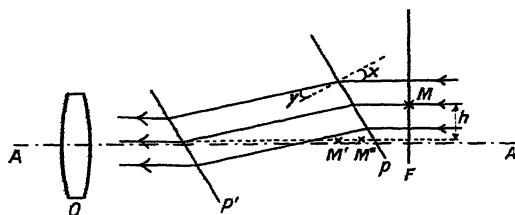
Figure 1.
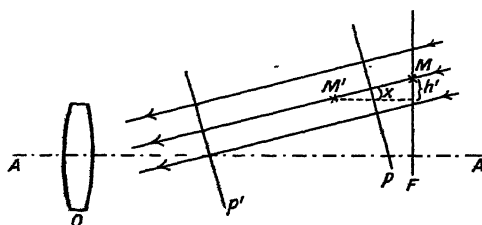
Figure 2.
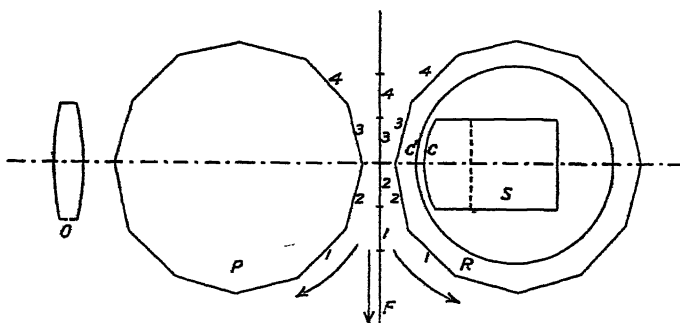
Figure 3.
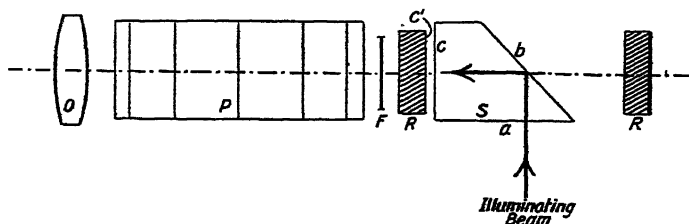
Figure 4.

light source a rotating optical device that influences the direction of the illuminating
beam without taking part directly in the formation of the projected image by the
objective.   Figures 3 and 4 show a schematic arrangement of the projector with a
suitable beam-deflecting device satisfying the optical condition just mentioned.
In figure 3, O is the projection objective, P the rotating polygonal prism which

produces a stationary image from the uniformly moving film F, R a rotating hollow polygonal prism which is the moving component of the beam-deflecting device. Inside R is the stationary component S with a cylindrical surface c concentric with the cylindrical inner surface c′ of the rotating component R. The uniform movement of the film F and the uniform rotation of P and R (as shown by the arrows) are mechanically synchronized and the consecutive pictures 1, 2, 3, 4 on the film correspond to the prism faces 1, 2, 3, 4 of P and R. The illuminating beam crosses the plane face a of the fixed prism S, the diagonal face b of which is a reflecting plane surface. The beam thus reflected at right angles crosses the concentric surfaces c and c′ and leaves the moving faces (2, 3 in figure 3) of the rotating polygonal prism R for traversing the film F, the rotating prism P and the objective O. The illuminating beam is more or less deflected according to the varying angle of refraction at the prism faces of the rotating polygon R. It is easily seen that the ratio of the angle of incidence of the principal ray traversing a point of the film to the angle between the prism face normal of P and the objective axis is approximately constant and less than 1. If this ratio has a suitable value, the non-linearity of the image shift is eliminated.

§3. FURTHER ABERRATIONS

In what follows, other aberrations produced by the rotating prism P will be discussed which, although less striking than the non-linearity of the image shift, must nevertheless be considered in any practical solution of the problem. These aberrations can be classified in two groups: (*a*) those which are essentially independent of the angular aperture of the illuminating beam, and (*b*) those which are proportional to the first and second powers of this angular aperture.

(*a*) The non-linearity of the image shift belongs to those aberrations which are essentially independent of the aperture; another is the chromatic variation of the image shift, which does not exceed the practically tolerable limit. A further aberration may arise as a consequence of film shrinkage, in which case the length of the film per image would become less than that on which the design of the prism is based. An optical compensation of film shrinkage by an automatically responding device is of course desirable when using considerably shrunk films in non-intermittent projectors; however, it would be difficult to discuss this question before treating the optical problems of the image formation in the standard case of non-shrunk film.

(*b*) The illuminating rays crossing the film have wide apertures in practical cases; consequently, the aberrations depending on the aperture need detailed investigation. These aberrations, which determine the quality of the projection if linearity of the image shift is achieved, will be treated in the following discussion of the image formation by wide-angle pencils starting from any point of the film.

Lens calculations, based on the approximation $\sin \alpha = \alpha - \alpha^3/6$, give account of the following five aberrations (apart from the chromatic error): spherical aberration, coma, astigmatism, curvature of field and distortion. It is possible to design the projection objective and the prism behind it, in the position in which the optic axis is perpendicular to the prism faces, together as a well-corrected lens system. Then the rotation of the prism produces periodic aberrations which have no axial symmetry and are to be related to the tangential and sagittal sections

of the rotating prism. These prismatic aberrations are obviously analogous to the five lens aberrations above mentioned. For instance, the non-linearity of the image shift is an effect corresponding to the distortion in lens optics. There is a prismatic astigmatism, while the field curvature of lens optics corresponds here to the varying positions of the plane of the astigmatic image. Finally, the periodically varying prismatic coma is also a very interesting and practically important aberration. The analogy between these periodic aberrations due to the rotating prism and the axially symmetrical aberrations of lens optics becomes immediately obvious by considering the variation of the former during the rotation of the prism. Thus the axially symmetrical aberrations due to the prism, when its two faces are just perpendicular to the optic axis, will change continuously to the tangential and sagittal prismatic aberrations.

§4. TANGENTIAL RAY-TRACE

We begin by tracing tangential rays through the prism. Thus figure 5 is crossing the prism axis at right angles. Let M be a point on the film; p, p′, two parallel faces of the prism; x, the angle between the objective axis and the prism face normal; e–e, the second principal plane of the objective; MO, a ray through M parallel to the objective axis, crossing the principal plane e–e at O after traversing the prism faces p, p′. Let further B₊ and B₋ be points of intersection of the plane e–e with two other rays from M; $+\tau$ and $-\tau$, the angle between these rays and MO; M′, virtual intersection of the rays MB₊ and MB₋; M′₊, virtual
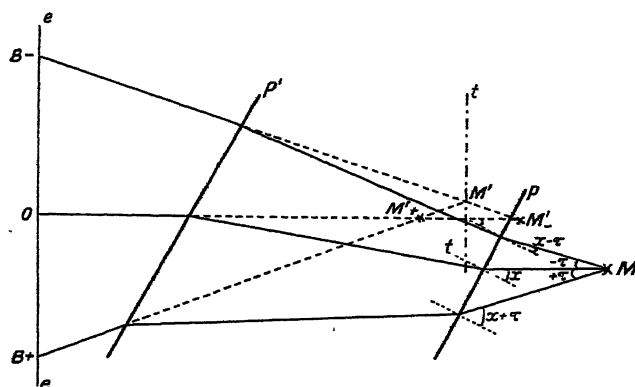
Figure 5.

intersection of the rays MO and MB₊; M′₋, virtual intersection of the rays MO and MB₋; t–t, a plane through M′ parallel to e–e and to the film plane.

We calculate first M′O₊ and M′₋O, i.e. the distances of these virtual image points from the principal plane. Let m be the distance of M from e–e. The parallel shift of the ray MO is

$$D \sin x \left(1 - \frac{\cos x}{\sqrt{(n^2 - \sin^2 x)}} \right) \qquad \ldots\ldots(4)$$

and the parallel shift of MB₊

$$D \sin (x+\tau) \left[1 - \frac{\cos (x+\tau)}{\sqrt{(n^2 - \sin^2 (x+\tau))}} \right], \qquad \ldots\ldots(5)$$

wherein D is again the diameter of the inscribed cylinder in the prism. Then the distance OB_+ is

$$m\tan\tau + D\sin x\left(1 - \frac{\cos x}{\sqrt{(n^2-\sin^2 x)}}\right) - D\sin(x+\tau)\left[1 - \frac{\cos(x+\tau)}{\sqrt{(n^2-\sin^2(x+\tau))}}\right]\frac{1}{\cos\tau}.$$

$$\dots\dots(6)$$

Introducing $t_+ = OM'_+$, we obtain

$$t_+ = \frac{OA_+}{\tan\tau} = m + D\frac{\sin x}{\tan\tau}\left(1 - \frac{\cos x}{\sqrt{(n^2-\sin^2 x)}}\right) - D\frac{\sin(x+\tau)}{\sin\tau}\left[1 - \frac{\cos(x+\tau)}{\sqrt{(n^2-\sin^2(x+\tau))}}\right],$$

$$\dots\dots(7)$$

which, breaking off the series of the trigonometric functions after the second power, results in

$$t_+ = m - D\frac{n-1}{n}\left[1 + \left(\frac{3}{2}\cdot\frac{n+1}{n^2} - \frac{1}{2}\right)x^2 + \frac{3}{2}\cdot\frac{n+1}{n^2}x\tau + \frac{n+1}{2n^2}\tau^2\right]. \quad\dots\dots(8)$$

Equation (8) reveals several aberrations:

The term in τ^2, being independent of the rotation of the prism, is simply the spherical aberration, and is the same as in the position when the prism faces are at right angles to the optic axis. The term in x^2 represents the position of the image formed by an infinitesimal pencil of rays in the tangential section through the prism. Finally, the term in $x\tau$ is a measure of the varying prismatic coma. It is of practical importance to find the intersection M' of the rays MB_+ and MB_-. The distance of M' from e–e is

$$m - D\frac{n-1}{n}\left[1 + \left(\frac{3}{2}\cdot\frac{n+1}{n^2} - \frac{1}{2}\right)x^2 + \frac{n+1}{2n^2}\tau^2\right] \quad\dots\dots(9)$$

and the distance of M' from the film plane

$$D\frac{n-1}{n}\left[1 + \left(\frac{3}{2}\cdot\frac{n+1}{n^2} - \frac{1}{2}\right)x^2 + \frac{n+1}{2n^2}\tau^2\right]. \quad\dots\dots(10)$$

The term in τ^2 is of no importance, because the spherical aberration of the prism (in the perpendicular position to the optic axis) can be corrected by the projection objective.

In the image plane t–t the tangential coma produces the vertical aberration

$$\frac{\overline{M'_+M'_-}}{2}\cdot\tan\tau = D\frac{n-1}{n}\cdot\frac{3}{2}\cdot\frac{n+1}{n^2}x\tau\cdot\tan\tau \approx D\cdot\frac{3}{2}\frac{n^2-1}{n^3}x\tau^2.$$

$$\dots\dots(11)$$

§5. SAGITTAL RAY-TRACE

The tracing of sagittal rays through the prism can be effected by reference to the perspective drawing figure 6. Taking M, a point on the film; MK, a ray parallel to the optic axis; MK', another ray in a sagittal plane, i.e. in a plane parallel to the prism axis (and parallel to the objective axis, in the particular case of figure 6) ; and taking also K and K', the points of incidence of these rays on the prism face p; V, the point of intersection of the perpendicular from M with the prism face, we get T for the point on this perpendicular for which MK⊥TK

and write x, the angle of incidence of the ray MK; α, the angle of incidence of the ray MK'; σ, the angle between MK and MK'; ϑ, the angle between MK' and K'T; ω, the angle between KT and K'T. We use the sign $\triangleleft$ for a right angle.

For the triangle KVK' we have

$$\sin^2\alpha = \sin^2\sigma + \cos^2\sigma \cdot \sin^2 x. \qquad \ldots\ldots(12)$$

It may be noted that the refracted rays of MK and MK'—which are not shown in figure 6—have no vertical intersection, unless σ is infinitely small. Consequently, there is no Pythagorian equation for the corresponding angles of refraction.
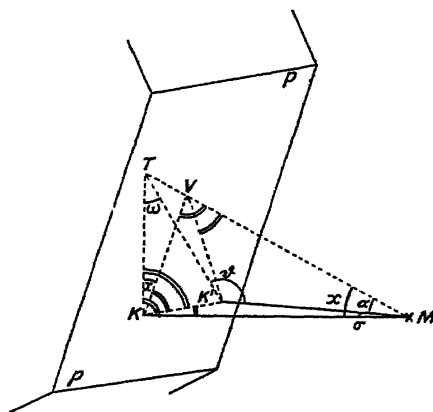
Figure 6.

We find for triangle KTK'

$$\frac{\sin\sigma}{\sin\omega} = \frac{\cos\sigma\cdot\tan x}{\cos\omega} \qquad \ldots\ldots(13)$$

and for triangle MK'T

$$\frac{\sin\sigma}{\sin\omega} : \frac{\cos\sigma}{\cos x} = \sin\alpha : \sin\vartheta. \qquad \ldots\ldots(14)$$

(13) and (14) give

$$\frac{\sin x}{\cos\omega} = \frac{\sin\alpha}{\sin\vartheta}. \qquad \ldots\ldots(15)$$

The ray MK', after refraction at two parallel faces of the prism, has suffered a parallel shift of magnitude,

$$d = D\sin\alpha\left(1 - \frac{\cos\alpha}{\sqrt{(n^2 - \sin^2\alpha)}}\right), \qquad \ldots\ldots(16)$$

in the plane of incidence. Produced backwards, the refracted ray intersects the film in a point whose distance from M is

$$d/\sin\vartheta, \qquad \ldots\ldots(17)$$

since the film plane and the triangle KTK' are parallel. The displacement (17) in the film plane has a vertical component

$$d\cos\omega/\sin\vartheta \qquad \ldots\ldots(18)$$

and a horizontal component

$$d \sin \omega / \sin \vartheta \qquad \dots\dots(19)$$

(the prism axis being assumed horizontal).

The vertical component (18) can be obtained from (15) and (16); if the power series for the trigonometrical functions are broken off after the second term, we have

$$d \frac{\cos \omega}{\sin \vartheta} = D \sin x \left(1 - \frac{\cos \alpha}{\sqrt{(n^2 - \sin^2 \alpha)}}\right) \approx D \frac{n-1}{n} x \left[1 + \left(\frac{n+1}{2n^2} - \frac{1}{6}\right) x^2 + \frac{n+1}{2n^2} \sigma^2\right].$$

$$\dots\dots(20)$$

The horizontal component (19) is given by (13) and (15)

$$d \frac{\sin \omega}{\sin \vartheta} = D . \tan \sigma . \cos x \left(1 - \frac{\cos \alpha}{\sqrt{(n^2 - \sin^2 \alpha)}}\right). \qquad \dots\dots(21)$$

Let us now consider in the plane MK'K of figure 6 another ray at an angle $-\sigma$ with MK, so that this new ray and the ray MK' have a point of intersection the distance of which from the film plane is

$$d \frac{\sin \omega}{\sin \vartheta} . \frac{1}{\tan \sigma} = D \cos x \left(1 - \frac{\cos \alpha}{\sqrt{(n^2 - \sin^2 \alpha)}}\right) \approx D \frac{n-1}{n} \left[1 + \left(\frac{n+1}{2n^2} - \frac{1}{2}\right) x^2 + \frac{n+1}{2n^2} \sigma^2\right].$$

$$\dots\dots(22)$$

In (22), the term in σ^2 is the spherical aberration of the prism, which does not change during the rotation, while the term in x^2 gives the position of the image formed by an infinitesimal pencil of rays in the sagittal section of the prism.

The difference between the x^2-terms of (10) and (22),

$$D \frac{n^2 - 1}{n^3} x^2, \qquad \dots\dots(23)$$

is the distance between the tangential and the sagittal points of intersection, i.e. the prismatic astigmatism.

In (20) the term with x^3 is the well known non-linear aberration of the parallel shift, and the term in $x\sigma^2$ is the sagittal coma.

§6. CONCLUSION

These aberrations have been calculated for the special example where the principal ray illuminating any point of the film is parallel to the objective axis. The formulae are the same for any other direction of the rays; in the general case, however, x does not mean the angle between the prism face normal and the objective axis, but the angle of incidence of the principal ray from any point on the film. τ and σ are always the tangential and sagittal angular semi-apertures of the wide-angle pencil illuminating a point of the film. In the general case, these aberrations as vector quantities have to be related to the direction of the principal ray in question, not to the axis of the objective.

A more detailed discussion of how the results given in the present paper can be used for the optical correction of the image formation in the rotating prism projector will be published later.

REFERENCE

TAYLOR, DENNIS H., 1937. *Proc. Phys. Soc.*, **49**, 663.

THE ROTATIONAL ANALYSIS OF SOME BANDS OF THE NEAR ULTRA-VIOLET SYSTEM (D–X) OF SILICON MONOSULPHIDE

BY R. F. BARROW,

Physical Chemistry Laboratory, University of Oxford

ABSTRACT. A rotational analysis has been done of the 0,1, 0,2, 0,3, 0,4, and 1,5 bands of the near ultra-violet band-system ($\nu_e \simeq 35030$ cm.$^{-1}$) of SiS. The transition responsible for the production of the band-system is of the type $D\,^1\Pi - X\,^1\Sigma$. The rotational constants have been derived from two sets of experimental observations : (i) of the bands in absorption, by G. Herzberg, N. Olson and F. J. Wilson (University of Saskatchewan), and (ii) of the bands in emission, by the author. The more important constants are given by

$$B_v'' = 0 \cdot 3036 - 0 \cdot 0014_9\,(v'' + \tfrac{1}{2}), \text{ for the ground state, and}$$
$$B_v' = 0 \cdot 2664 - 0 \cdot 0020_5\,(v' + \tfrac{1}{2}) \text{ for the upper state.}$$

Λ-type doubling in the upper $^1\Pi$ state is negligibly small, at least up to $J = 100$. The values of the equilibrium internuclear distances are $r_e'' = 1 \cdot 929$ A. and $r_e' = 2 \cdot 059_5$ A. A brief discussion of the results is given.

---

§1. INTRODUCTION

THE production and vibrational analysis of a near ultra-violet band-system of silicon monosulphide ($\nu_e \simeq 35030$ cm.$^{-1}$) were described some time ago by Barrow and Jevons (1938). Since then, a preliminary account has been given of the rotational analysis of the 0,1, 0,2 and 0,3 bands of this system (Barrow, 1944), which was shown to be due to a transition of the type $^1\Pi - {}^1\Sigma$. The present paper gives the results of a more detailed analysis of these three bands and of the 0,4 and 1,5 bands of the same system.

Two distinct sets of experimental material have been used. When the preliminary analysis of the 0,1, 0,2 and 0,3 bands from my own plates, taken in emission, had been completed, I learned from Professor G. Herzberg that Mr. N. Olson and Mr. F. J. Wilson, working in his laboratory at Saskatchewan on plates of the bands in absorption taken by Olson, had already carried out analyses of the 0,1, 0,2, 0,3, 0,4 and 0,5 bands. Their results proved to be in close agreement with mine. Professor Herzberg then most kindly sent me the detailed measurements and analyses: the results to be given are based, where possible, on the averaged measurements of the two independent sets of investigations.

§2. EXPERIMENTAL METHODS

Olson photographed the bands in absorption. SiS, prepared by the method described by Barrow and Jevons (1938), was sublimed into an electrically-heated T-shaped quartz tube. This consisted of an absorption tube, 15–20 cm. long and 2·5 cm. in internal diameter, fitted with a central T-piece which acted as a reservoir for SiS and through which the whole system could be evacuated and

ultimately sealed. Independent heating of the reservoir and of the main tube minimized deposition of SiS on the windows of the latter. The tubes were operated at about 900° c. · A hydrogen lamp was used as source of continuous radiation, and the bands were photographed in a second order of a 6-m. concave grating spectrograph on Eastman " 40 " plates.

The experimental arrangements used in developing the SiS system in emission have already been described (Barrow and Jevons, 1938). The region 2800–3100 A. was photographed in a fourth order of the 10-ft. concave grating at Imperial College, London. The grating itself largely absorbed the overlapping fifth order: the overlapping part of the third order was absorbed in a 2-cm. column of an aqueous solution of cobalt and nickel sulphates containing 4·15 g. $CoSO_4.7H_2O$ and 14·5 g. $NiSO_4.H_2O$ dissolved in 100 c.c. water (Bowen, 1942). Ilford " Zenith " plates were used, and exposure times were of the order of two hours.

§ 3. ROTATIONAL ANALYSIS

The 0,1, 0,2, and 0,3 bands were first selected for analysis as their structure is little overlaid by that of other bands. However, the 0,2 band, which appeared to be particularly simple, proved to suffer from considerable overlapping of the branches, and the 0,1 and 0,3 bands were therefore examined first. In no bands were the lines resolved close to the heads, but further away from the heads three branches were apparent, one of which—the Q branch—was considerably stronger than either of the others.

By a trial and error study of the first differences, $R(J) - Q(J) \simeq Q(J+1) - P(J+1)$, it turned out to be an easy matter, first to identify the P and R branches, then to ascertain the absolute J-numbering, and finally, using a rough value of D, obtained from the relation $D = -4B^3/\omega$, to get fairly accurate values of the constants B for the 0,1, 0,3 and 0,2 bands.

At this stage I received the data of Olson on the 0,2 and 0,3 bands and of Wilson on the 0,1, 0,4, and 1,5 bands. Their assignments of the lines of the 0,1, 0,2 and 0,3 bands agreed with mine in detail, and, after taking account of a small systematic difference in the wave-numbers of the lines (my values were on an average 0·35 $cm.^{-1}$ lower), means of the wave-numbers of the lines of these three bands were taken. Only Wilson measured the 0,4 and 1,5 bands. The wave-numbers are displayed in table 1.

The procedure adopted for the final evaluation of the constants was as follows. The constants which could be determined with the highest absolute accuracy were obviously those for the upper state with $v' = 0$, for which there were measurements of four bands. Either $\Delta_1 F'(J)$ or $\Delta_2 F'(J)$ differences could have been used: eventually the former,

$$\Delta_1 F'(J) = R(J) - Q(J) \simeq Q(J+1) - P(J+1),$$

were chosen after it had been demonstrated that, within the limits of error of the measurements, the differences $R(J) - Q(J)$ were equal to $Q(J+1) - P(J+1)$, up to the highest J-values ($J \sim 100$). Thus the Λ-type doubling of the upper $^1\Pi$ state is very small—as would be expected by analogy with the results of the analysis of bands of the similar system of SiO (Saper, 1932).

Table 1 Observed wave numbers of rotational lines (cm.⁻¹, *in vac.*)

Asterisks indicate unresolved lines

v', v''	0,1			0,2			0,3			0,4			1,5		
J	Q(J)	R(J)	P(J)	Q(J)	R(J)	P(J)	Q(J)	R(J)	P(J)	Q(J)	R(J)	P(J)	Q(J)	R(J)	P(J)
20	34148·60*	34159·76*	34138·36*	33409·88*		33399·89*	32676·16*	32687·48*	32665·68*	31948·17*		31937·13*	31730·97*		31720·39*
21	47·36*	58·94*	36·32*	08·36*		97·93*	74·97	86·76*	63·54*	46·86*		35·34*	29·51*		18·71*
22	45·52*	57·73*	34·34*	06·79*		95·74*	73 46	85·81*	61·63*	45·50*		33·18*	28·08*		16·75*
23	43·79*	56·68*	31·71	05·25*		93·42*	72·15	84·76*	59·57*	43·84*		31·76*	26·90*		14·73*
24	42·13*	55·36*	29·55	03·45*		91·20*	70·55	83·52*	57·52*	42·27		29·74*	25·33		12·79*
25	40·31*	54·07*	27·12	01·71*		88·97*	68·80	82 37*	55·39*	40·58*		27·62*	23·61*		10·69*
26	38·36*	52·87*	24·68	399·89*		86·68*	67·10	81·20*	53·11*	38·90*		25·53*	22·30*		08·43*
27	36·32*	51·57*	22·10	97·93*		84·34*	65·25	80·00*	50·83*	37·13*		23·28*	20·39*	31734·85*	06·18*
28	34·34*	50·11*	19·63	96·24		81·71*	63·54*	78·88*	48·44*	35·34*	31950·40*	21·40*	18·71*	33·52*	03·91*
29	32·51	48·60*	17·07	94·32		79·28*	61·63*	77·60*	45·99*	33·18*	49·54*	18·69*	16·75*	32·33*	01·59*
30	30·08	46·90	14·46	92·12		76·79*	59·57*	75·90*	43·49*	31·76*	48·17*	15·84	14·73*	30·97*	699·31*
31	27·81	44·96	11·65	90·05		73·98*	57·52*	74·38*	40·95*	29·74*	46·86*	13·44	12·79*	29·51*	96·51*
32	25·89*	43·31	08·77	87·76	33405·25*	71·25*	55·39*	72·72*	38·41*	27·62*	45·50*	10·83	10·69*	28·08*	94·01*
33	23·42*	41·38	05·91	85·55	03·45*	68·36*	53·11*	71·11*	35·75*	25·53*	43·84*	08·23	08·43*	26·90*	91·30*
34	20·86*	39·43	02·83	83·17	01·71*	65·48*	50·83*	69·29*	32·94*	23·28*	42·27*	05·54	06·18*	25·33*	88·62*
35	18·33*	37·38	099·77	80·61	399·89*	62·39*	48·44*	67·61*	30·20*	21·40*	40·58*	02·80	03·91*	22·30*	85·82*
36	15·84*	35·32	96·57	78·23	97·93*	59·39*	45·99*	65·68*	27·36*	18·69*	38·90*	899·98	01·59*	20·39*	83·35*
37	13·08*	32·86	93·42	75·62	95·74*	56·34*	43·49*	63·54*		16·91*	37·13*	97·10	699·31*	18·71*	80·48*
38	10·25*	30·84	90·21	73·04	93·42*	53·14*	40·95*	61·63*		14·43*	35·34*	94·21	96·51*	16·75*	77·09*
39	07·36*	28·43	86·74	70·38	91·20*	49·77*	38·41*	59·57*		12·07*	33·18*	91·08	94·01*	14·73*	73·80*
40	04·41*	25·89*	83·31	67·41	88·97*	46·35*	35·75*	57·52*		09·39*	30·85	88·09	91·30*	12·79*	70·63*
41	01·45*	23·42*	79·87	64·63	86·68*	43·06*	32·94*	55·39*	11·59	06·80*	28·76	85·00	88·62*	10·69*	67·49*
42	098·29*	20·86*	76·31	61·92	84·34*	39·57*	30·20*	53·11*	08·20	04·07*	26·65	81·74	85·82*	08·43*	64·15*
43	95·13*	18·33*	72·52	58·84	81·71*	36·09*	27·36*	50·83*	04·79	01·28*	24·42	78·47	83·35*	06·18*	61·05*
44	91·89	15·84*	68·75	55·88	79·28*	32·41*	24·46	48·44*	01·19	898·43*	22·18	75·12	80·48*	03·91*	57·36*
45	88·64*	13·08*	64·96	52·74	76·79*	28·81*	21·35	45·99*	597·74	95·56*	20·10	71·76	77·09*	01·59*	53·86*
46	85·33*	10·25*	61·17	49·77*	73·98*	24·91*	18·30	43·49*	94·26	92·57*	16·91*	68·33	74·65	699·31*	50·51*
47	81·78*	07·36*	57·45	46·35*	71·25*	21·11*	15·15	40·65	90·34	89·50*	14·43*	64·77	71·65	96·51*	46·69*
48	78·35*	04·41*	53·23	43·06*	68·36*	17·24*	12·01	37·90	86·61	86·46*	12·07*	61·10	68·38	94·01*	43·11*
49	74·92	01·45*	49·20	39·57*	65·48*	13·37*	08·77	35·20	82·93	83·26*	09·39*	57·51	64·92	91·30*	39·45*

50	71·38	098·29*	44·87	36·09*	62·39*	09·34*	05·40	32·40	78·95	79·99*	06·80*	53·76	62·70	88·62*	35·46*
51	67·63	95·13*	40·63	32·41*	59·39*	05·22*	01·97	29·60	75·02	76·80*	04·07*	49·91	59·32	85·82*	31·66*
52	63·80	91·89*	36·48	28·81*	56·34*	01·01*	598·49	26·55	71·05	73·27*	01·28*	45·97	55·66	83·35*	27·79*
53	59·91	88·64*	31·95	24·91*	53·14*	296·78*	94·92	23·45	66·97	69·95*	898·43*	42·08	51·85	80·48*	23·49*
54	56·02	85·33*	27·51	21·11*	49·77*	92·48*	91·31	20·25	62·78	66·43*	95·56*	38·18	48·18	77·09*	19·35*
55	51·94	81·78*	22·94	17·24*	46·35*	88·07*	87·62	17·20	58·60	62·89*	92·57*	34·06	44·66	73·80*	16·25
56	47·78	78·35*	18·51	13·37*	43·06*	83·66*	83·81	13·91	54·30	59·55	89·50*	29·95	40·93	70·63*	12·02
57	43·58	74·29	13·58	09·34*	39·57*	79·24*	79·98	10·58	50·00	55·94	86·46*	25·88	37·14	67·49*	07·44
58	39·42	70·55	08·99	05·22*	36·09*	74·69*	76·12	07·10	45·66	52·13	83·26*	21·55	33·26	64·15*	03·23
59	35·12	66·93	04·03	01·01*	32·41*	69·94*	72·16	03·67	41·01	48·30	79·99*	17·13	29·55	61·05*	598·63
60	30·66	62·96	33999·07	296·78*	28·81*	65·27*	68·05	00·25	36·49	44·41	76·80*	12·85	25·60	57·36*	94·18
61	26·26	58·84	94·12	92·48*	24·91*	60·47*	63·94	596·57	31·78	40·48	73·27*	08·35	21·54	53·86*	89·69
62	21·69	54·81	88·92	88·07*	21·11*	55·58*	59·77	93·07	27·20	36·45	69·95*	04·18*	17·49	50·51*	85·18
63	17·01	50·70	83·95	83·66*	17·24*	50·61*	55·51	88·94	22·40	32·47	66·43*	799·65*	13·27	46·69*	80·54
64	12·33	46·52	78·63	79·24*	13·37*	45·59*	51·18	85·29	17·62	28·28	62·89*	94·80*	09·05	43·11*	75·73
65	07·53	42·15	73·24	74·69*	09·34*	40·41*	46·76	81·47	12·56	23·99	58·73	90·05*	04·83	39·45*	70·94
66	02·53	37·83	67·92	69·94*	05·22*	35·26*	42·17	77·57	07·59	19·75	54·92	85·25*	00·48	35·46*	66·11
67	33997·83	33·49	62·50	65·27*	01·01*	30·01*	37·76	73·58	02·60	15·50	51·13	80·42*	595·85	31·66*	61·07
68	92·74	29·12	57·07	60·47*	296·78*	24·65*	33·17	69·39	497·43	11·08	47·26	75·43*	91·42	27·79*	56·03
69	87·68	24·44	51·39	55·58*	92·48*	19·20*	28·44	65·28	92·26	06·63	43·29	70·36*	86·99	23·49*	50·82
70	82·51	19·84	45·99*	50·61*	88·07*	13·74*	23·71	61·01	86·94	02·07	39·28	65·29*	82·27	19·35*	45·91
71	77·22	15·05	40·23*	45·59*	83·66*	08·19*	19·04	56·74	81·66	797·41	35·21	60·20*	77·58	14·99	40·66
72	71·91	10·21	34·38*	40·41*	78·59	02·63*	14·09	52·45	76·25	92·80	31·09	54·86*	72·82	10·66	35·47
73	66·62	05·45	28·45*	35·26*	74·09	196·89*	09·09	47·90	70·78	87·98	26·94	49·57*	67·92	06·46	29·62
74	61·04	00·46	22·49*	30·01*	69·29	91·15*	04·07	43·29	65·30	83·14	22·59	44·20*	63·02	02·09	24·58
75	55·58	33995·23	16·31*	24·65*	64·45	85·27*	498·97	38·72	59·50	78·24	18·26	38·47	58·02	597·56	
76	49·88	90·23	10·22*	19·20*	59·59	79·34*	93·75	34·05		73·24	13·67		53·00	92·92	
77	44·33	85·00	03·87*	13·74*	54·67	73·28*	88·50	29·20		68·22	09·12		47·86	88·39	
78	38·57	79·73	897·49*	08·19*	49·55	67·20*	83·20	24·35		63·27	04·18*		42·58	83·86	
79	32·52	74·25	91·05*	02·63*	44·38	61·11*	77·65	19·45		58·06	799·65*		37·47	78·94	
80	26·55	68·91	84·53*	196·89*	39·12	54·83*	72·15	14·65		52·80	94·80*		31·94	74·03	
81	20·58	63·42	77·95*	91·15*	34·09	48·47*	66·60			46·96	90·05*		26·68	69·42	
82	14·40	57·61	71·51	85·27*	28·66	42·20*	60·90			42·08	85·25*		21·24	64·31	
83	08·07	51·86	64·97	79·34*	23·17	35·68*	55·35			36·49	80·42*			59·19	
84	01·96	45·99*	58·31	73·28*	17·71	29·17*	49·60				75·43*				

Table 1—*continued*

Asterisks indicate unresolved lines

v', v''	0,1			0,2			0,3			0,4			1,5		
J	$Q(J)$	$R(J)$	$P(J)$	$Q(J)$	$R(J)$	$P(J)$	$Q(J)$	$R(J)$	$P(J)$	$Q(J)$	$R(J)$	$P(J)$	$Q(J)$	$R(J)$	$P(J)$
85	895·61	40·23*	51·29	67·20*	12·17	22·55*	43·75				70·36*				
86	89·28	34·38*	44·36	61·11*	06·39	15·89					65·29*				
87	82·62	28·45*	37·37	54·83*	00·69	09·24					60·20*				
88	76·14	22·49*	30·29	48·47*	194·76	02·44					54·86*				
89	69·42	16·31*	23·05	42·20*	88·99	095·54					49·57*				
90	62·74	10·22*	15·82	35·68*	82·88	88·39					44·20*				
91	55·78	03·87*	08·37	29·17*	76·78	81·34					38·47*				
92	48·87	897·47*	00·93	22·55*	70·76	74·29									
93	41·94	91·05*	793·53	15·59	64·56	66·94									
94	34·86	84·53*	85·98	08·79	58·12	59·74									
95	27·91	77·95*	78·10	01·89	51·85	52·64									
96	20·35	71·17	70·35	094·99	45·33	45·09									
97	13·19	64·16	62·86	87·89	38·91	37·54									
98	05·67	57·15	54·74	80·74	32·16	29·84									
99	798·24	50·17	47·20	73·54	25·41	22·04									
100	90·65	43·16	38·85	66·29	18·74	14·44									
101	83·34	35·95	30·91	59·09	11·88	06·79									
102	75·48	28·89	22·43	51·64	04·79	32998·64									
103	67·67		14·13	43·99	097·84	90·69									
104	59·71		05·82	36·34	90·79	82·64									
105	51·50			28·74	83·59										
106	43·49			20·99	76·34										
107	35·32			12·99	69·14										
108	27·02			05·19	61·59										
109	18·64			32997·24	54·04										
110	10·47			89·04	46·74										
111	02·06			80·99	38·99										

It is a familiar result that

$$\frac{\Delta_1 F(J)}{(J+1)} = 2B + 4D(J+1)^2 + 6F(J+1)^4 + \ldots, \qquad \ldots\ldots(1)$$

where B, D and F are the constants in the expression for the rotational energy levels,

$$F(J) = B \cdot J(J+1) + D \cdot J^2(J+1)^2 + F \cdot J^3(J+1)^3 + \ldots, \qquad \ldots\ldots(2)$$

and generally, $B \gg D \gg F$.

Thus, a plot of $\dfrac{\Delta_1 F(J)}{J+1}$ against $(J+1)^2$ should—at least for moderate values of J—yield a straight line of intercept $2B$ and of slope $4D$. A selection of $\Delta_1 F_0'(J)$ values for $v' = 0$ is given in table 2, and a plot of $\dfrac{\Delta_1 F_0'(J)}{J+1}$ against
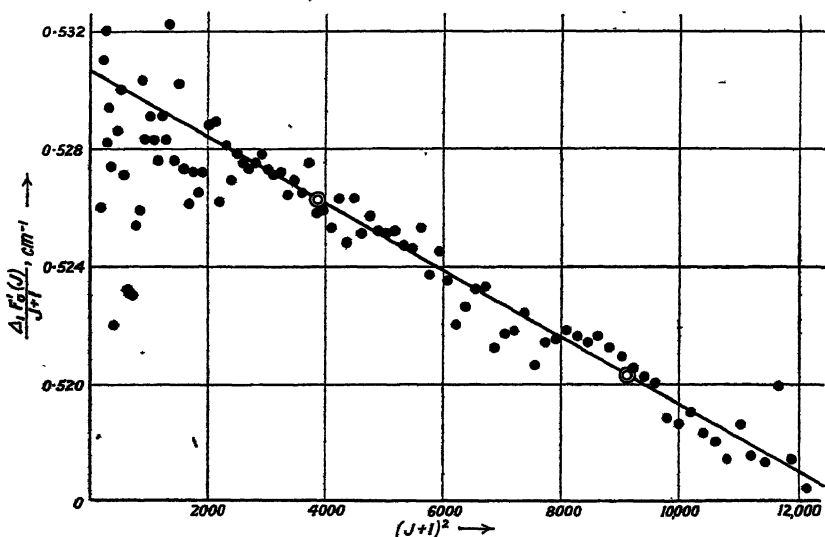
Figure 1. $\dfrac{\Delta_1 F_0'(J)}{J+1}$ as a function of $(J+1)^2$.

The two points ⊚ represent the mean values for $J = 45$–76, 77–111 : their coordinates are $(0.52626, 3868)$ and $(0.52027, 9127)$, giving $2B_0' = 0.5306_5$, $4D_0' = 1.14 \times 10^{-6}$.

$(J+1)^2$ is shown in figure 1. From the latter it appeared safe—in the first instance—to neglect terms in F, and the best straight line was therefore drawn through all the points with $J > 45$, using the centre-of-gravity method. This gave $B_0' = 0.2653_3$ and $D_0' = -2.85 \times 10^{-7}$ cm^{-1}.

It was considered that the measurements of the Q-branch lines were more reliable than those of either P- or R-branch lines, and the lower-state constants for the 0,1, 0,2, 0,3, and 0,4 bands were therefore determined from the Q line in the following way. The wave-number of any such line is given by

$$Q(J) = v_0 + (B' - B'') \cdot J(J+1) + (D' - D'') \cdot J^2(J+1)^2 + \ldots, \qquad \ldots\ldots(3)$$

where v_0 is the band-origin. Provisional values of B'' and of D'', obtained from $\Delta_1 F''(J)$ and $\Delta_2 F''(J)$ differences, were inserted in (3), and preliminary values of

ν_0, which we may write $\bar{\nu}_0$, calculated for each line, thus

$$\bar{\nu}_0 = Q(J) - \overline{(B'-B'')}.J(J+1) - \overline{(D'-D'')}.J^2(J+1)^2 - \ldots, \quad \ldots\ldots(3\,\mathrm{a})$$

where the horizontal bars indicate provisional values. In general these $\bar{\nu}_0$ values were not quite constant with J, since the $\overline{(B'-B'')}$ and $\overline{(D'-D'')}$ values were not quite correct. If the differences between the true and the provisional values are indicated by Δ's—thus, $\Delta\nu_0 = \nu_0 - \bar{\nu}_0$,—then

$$\Delta\nu_0 = -\Delta(B'-B'').J(J+1) - \Delta(D'-D'').J^2(J+1)^2 - \ldots \quad \ldots\ldots(3\,\mathrm{b})$$

The $\bar{\nu}_0$ values were then plotted against $J(J+1)$: the intercept gave ν_0, and the slope at low values of $J(J+1)$ the correction $-\Delta(B'-B'')$. New values of $\bar{\nu}_0$

Table 2.   Some values of $\Delta_1 F_0'(J)$

J	$R(J)-Q(J)$				$Q(J+1)-P(J+1)$				Mean obs.	Calc.
	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4		
40	21·48*	21·56*	21·77*	21·46*	21·58*	21·57*	21·35*	21·80*	21·57	21·68
41	1·97*	2·05*	2·45*	1·96*	1·98*	2·35*	2·00*	2·33*	2·14	2·21
42	2·57*	2·42*	2·91*	2·58*	2·61*	2·75*	2·57*	2·81*	2·65	2·73
43	3·20*	2·87*	3·47*	3·14*	3·14*	3·47*	3·27	3·31*	3·24	3·25
44	3·95*	3·40*	3·98*	3·75*	3·68*	3·93*	3·61	3·80*	3·75	3·78
45	4·44*	4·05*	4·64*	4·54*	4·16*	4·86*	4·04	4·24*	4·33	4·30
65	34·62	34·65*	34·71	34·74	34·61	34·68*	34·58	34·50*	34·64	34·69
66	5·30	5·28*	5·40	5·17	5·33	5·26*	5·16	5·08*	5·26	5·21
67	5·66	5·74*	5·82	5·63	5·67	5·82*	5·74	5·65*	5·71	5·72
68	6·38	6·31*	6·22	6·18	6·29	6·38*	6·18	6·27*	6·27	6·25
69	6·76	6·90*	6·84	6·66	6·52*	6·87*	6·77	6·78*	6·76	6·76
70	7·33	7·46*	7·30	7·21	6·99*	7·40*	7·38	7·21*	7·29	7·27
90	47·48*	47·20*	—	—	47·41	47·83*	—	—	47·47	47·42
91	8·09*	7·61*	—	—	7·94	8·26*	—	—	7·97	7·92
92	8·60*	8·21*	—	—	8·41	8·65	—	—	8·49	8·42
93	9·11*	8·97	—	—	8·88	9·05	—	—	8·99	8·92
94	9·67*	9·33	—	—	9·81	9·25	—	—	9·49	9·42
95	50·04*	9·96	—	—	50·00	9·90	—	—	9·97	9·92

* Asterisks indicate values involving the wave-numbers of unresolved lines.

were then calculated, with the corrected $(B'-B'')$ terms, but with the same $\overline{(D'-D'')}$ terms. Plots of $\Delta\nu_0$ from these $\bar{\nu}_0$ values against $J^2(J+1)^2$ enabled $\Delta(D'-D'')$ to be evaluated.

When this had been done, it was found that the values of ν_0 calculated from (3) with the now corrected values of $(B'-B'')$ and of $(D'-D'')$ still showed some systematic departure from constancy in the region of the highest J values (90–110). The magnitude of terms in $(F'-F'').J^3(J+1)^3$ was therefore considered. F' and F'' were calculated from the relation

$$F = \frac{2D^2}{B} + \frac{\alpha_1\omega D^2}{6B^3}$$

(see, for example, Jevons, 1932). The results were

$$F' = -1.7 \times 10^{-13} \, \text{cm}^{-1},$$

$$F'' < 10^{-14} \, \text{cm}^{-1},$$

so that, approximately,

$$(F' - F'') = -1.7 \times 10^{-13}.$$

The addition of the terms $-1.7 \times 10^{-13} \cdot J^3(J+1)^3$ to (3) was found to lead to values of ν_0 which were then sensibly constant over the whole range of J values for all these four bands.

The constants derived in this way were finally checked by calculating the wave numbers of all the measured lines of the four bands, by first evaluating the Q branches and then using the calculated $\Delta_1 F_0'(J)$ values to get the P- and R-branch lines.

The lines of the 1,5 band were so often superimposed on other lines of the same band that accurate evaluation of the rotational constants from either $\Delta_1 F(J)$ or $\Delta_2 F(J)$ differences was found to be impossible. Accordingly, values of B_5'' and of D_5'' were obtained by linear extrapolation, and the quantities $(B_1' - B_5'')$ and $(D_1' - D_5'')$ determined from the Q-branch lines in the way described above.

§4. RESULTS AND DISCUSSION

The vibrational and rotational constants are collected together in tables 3 and 4: some comments follow.

(i) The structure of the bands makes it quite certain that the electronic transition is of the type $^1\Pi - ^1\Sigma$. The ground state of SiS is most probably $^1\Sigma^+$ (see, for example, Herzberg, 1939). The system is therefore exactly analogous to the "4th Positive" $A^1\Pi - X^1\Sigma^+$ system of carbon monoxide, and to the similar system of SiO, as was, indeed, suspected before the rotational analysis.

(ii) The energies and products of dissociation of SiS in these two states will be considered in a later paper on another ultra-violet system of SiS (see also Vago and Barrow, 1946). However, it may be noted that the fact that the $D^1\Pi$ state is a singlet makes it most probable that it, like the ground state, dissociates into the ground state atoms, $Si(^3P) + S(^3P)$.

(iii) The new values of ω_e'' and of $x_e''\omega_e''$, determined from the band-origins, are very close to the values obtained from the original band-head measurements. As the analysis shows, this agreement arises from the circumstances that the band-heads are not very far from the origins (~ 2 cm^{-1}), and that the separation, head-origin, changes only rather slowly from band to band.

(iv) It is to be expected that SiS should have a similar electronic configuration in its ground state to the isosteric molecule P_2, which also has 30 extra-nuclear electrons. As with other sets of such molecules, this may account for the rather striking similarity in the ground-state constants (table 5). However, no upper state of P_2 corresponding to the $D^1\Pi$ level of SiS is known.

(v) The bond distance in the ground state of SiS, about 1·93 A., is some 0·22 A. shorter than the uni-covalent distance (Schomaker and Stevenson, 1940). The bond in SiS is, of course, expected to be stronger, and thus shorter, than a

single covalent bond, and the observed degree of shortening is in reasonable accord with the figures of 0·25 and 0·15 A. for the ground states of the similar molecules SiO and GeO.

Table 3. Constants of the $D\,^1\Pi$ and $X\,^1\Sigma^+$ states of SiS

State :	$X\,^1\Sigma^+$		$D\,^1\Pi$	
ν_e	0	cm.$^{-1}$	35028·8 cm.$^{-1}$	H
ω_e	749·65		512·0	H
	(749·5)	H		
$x_e\omega_e$	−2·575		−2·38	H
	(−2·56)	H	−0·045	H
$y_e\omega_e$	—			
B_4	0·2969$_3$		—	
B_3	0·2984$_1$		—	
B_2	0·2999$_3$		—	
B_1	0·3014$_0$		0·2632$_8$	
B_0	0·3028$_9$		0·2653$_3$	
B_e	0·3036$_3$		0·2663$_6$	
a_1	−0·0014$_9$		−0·0020$_5$	
D_4	−1·9$_7\times10^{-7}$		—	
D_3	−1·9$_8\times10^{-7}$		—	
D_2	−1·9$_8\times10^{-7}$		—	
D_1	−2·0$_0\times10^{-7}$		−2·95$\times10^{-7}$	
D_0	—		−2·85$\times10^{-7}$	
F	—		−1·7$\times10^{-13}$	
r_0	1·931$_4$ A.		2·063$_6$ A.	
r_e	1·929$_0$ A.		2·059$_5$ A.	

Notes : (1) The values of r were obtained from $r=\dfrac{4\cdot1065}{\sqrt{\mu B}}$ A., where μ is the reduced mass in physical atomic weight units. μ was calculated for ^{28}Si ^{32}S, with Si=27·987 and S=31·982.

(2) H signifies a constant derived from band-head data.

Table 4. Positions of band-origins

Band	Origin
0,1	34163·8$_8$ cm.$^{-1}$
0,2	3424·5$_4$
0.3	2690·3$_1$
0,4	1961·2$_7$
1,5	1744·6$_1$

Table 5. Comparisons of the constants of SiS with those of the isosteric molecule P_2*

	State	ω_e	Force constant	Dissociation energy	B_e	r_e
SiS	$X\,^1\Sigma^+$	749·65	4·94×10^5 dyne/cm.	6·6 ev.	0·3036	1·929 A.
P_2	$X\,^1\Sigma_g^+$	780·43	5·56	5·03	0·3031	1·895

* Herzberg, Herzberg and Milne (1940).

§ 5. ACKNOWLEDGMENTS

It gives me great pleasure to record my thanks to Professor Herzberg, Mr. Olson and Mr. Wilson for their generosity in putting their results at my disposal and to Professor H. Dingle for facilities in the spectroscopic laboratories of the Department of Physics, Imperial College, London.

REFERENCES

BARROW, R. F., 1944. *Nature, Lond.*, 154, 364.
BARROW, R. F. and JEVONS, W., 1938. *Proc. Roy. Soc.*, A, 169, 45.
BOWEN, E. J., 1942. *Chemical Aspects of Light* (Oxford University Press), p. 179.
HERZBERG, G., 1939. *Molecular Spectra and Molecular Structure*: I. *Diatomic Molecules.* (New York: Prentice Hall), p. 368.
HERZBERG, G., HERZBERG, L. and MILNE, G. G., 1940. *Canad. J. Res.*, A, 18, 139.
JEVONS, W., 1932. *Report on Band Spectra of Diatomic Molecules* (Physical Society), p. 28.
SAPER, P. G., 1932. *Phys. Rev.*, 42, 498.
SCHOMAKER, V. and STEVENSON, D. P., 1941. *J. Amer. Chem. Soc.*, 63, 37.
VAGO, E. E. and BARROW, R. F., 1946 a. *Nature, Lond.*, 157, 77.
VAGO, E. E. and BARROW, R. F., 1946 b. *Proc. Phys. Soc.*, 58, 538.

VELOCITY OF SOUND IN MIXTURES OF ARGON, HELIUM AND HYDROGEN AT LOW TEMPERATURES

BY A. VAN ITTERBEEK AND W. VAN DONINCK,
Physical Laboratory, University of Louvain, Belgium

ABSTRACT. Measurements are described of the velocity of ultrasonics in mixtures of argon, helium and hydrogen at a frequency of 524 kc./sec. as a function of pressure and at fixed temperatures near 90° and 20° absolute. From these data, information about the interaction between the gases in the mixtures is deduced, using the equation of state of Kamerlingh Onnes.

---

§ 1. INTRODUCTION

SOME years ago one of us, with W. H. Keesom, developed an indirect method of determining the second virial coefficient of gases at low temperatures (v. Itterbeek and Keesom, 1930). This method is based on measurements of the velocity of sound as a function of pressure at different fixed temperatures. Since then this method has been employed in our laboratory using ultrasonics, and has been applied to the following gases: A, O_2, H_2, D_2, CO (v. Itterbeek *et al.*, 1938, 1943, 1944). In the last research we extended this method to gas mixtures. In this paper we report on measurements carried out on A–He, A–H_2 and H_2–He respectively at liquid-oxygen and liquid-hydrogen temperatures. From these measurements it is possible to obtain some information about the interaction between the gases concerned.

§ 2. METHOD

The velocity of sound W can be expressed by means of the equation

$$W^2 = -\frac{c_p}{c_v} \cdot \left(\frac{\partial p}{\partial v}\right)_T \cdot \frac{v^2}{M}. \qquad \ldots\ldots(1)$$

Taking into account the equation of state of Kamerlingh Onnes,

$$pv = RT\left(1 + \frac{B}{v}\right), \qquad \ldots\ldots(2)$$

equation (1) can be transformed into

$$W = W_0\left(1 + \frac{S}{R \cdot T} \cdot p\right), \qquad \ldots\ldots(3)$$

W_0 being the velocity of sound in the gas mixture in the ideal state ($p=0$).

$$W_0 = \frac{c_p \cdot x + c_p'(1-x)}{c_v \cdot x + c_v'(1-x)} \cdot \frac{R \cdot T}{M_1 \cdot x + M_2(1-x)} \qquad \ldots\ldots(4)$$

(M_1, M_2 are the molecular masses, R the gas constant, c_p, c_p', c_v and c_v' the specific heats for the ideal state).

Further,

$$S = B + \frac{T}{\lambda} \cdot \frac{dB}{dT} + \frac{T^2}{2\lambda(\lambda+1)} \cdot \frac{d^2B}{dT^2}. \qquad \ldots\ldots(5)$$

λ is a numerical factor given by

$$\lambda = \frac{x \cdot c_v + (1-x) \cdot c_v'}{M_1 \cdot x + (1-x) \cdot M_2}. \qquad \ldots\ldots(6)$$

For the case of He–A and He–H_2 respectively at liquid-oxygen and liquid-hydrogen temperatures, we have $\lambda = 3/2$.

Writing

$$B = B_1 \cdot x^2 + \beta x \cdot (1-x) + B_2 \cdot (1-x)^2 \qquad \ldots\ldots(7)$$

together with equation (5), we obtain

$$S = S_1 \cdot x^2 + \alpha x \cdot (1-x) + S_2 \cdot (1-x)^2 \qquad \ldots\ldots(8)$$

with

$$\alpha = \beta + \frac{T}{\lambda} \cdot \frac{d\beta}{dT} + \frac{T^2}{2\lambda(\lambda+1)} \cdot \frac{d^2\beta}{dT^2}. \qquad \ldots\ldots(9)$$

For the case where the two components do not possess the same specific heats (A–H_2), λ depends on concentration (6). Equation (8) must be changed because the factors S_1 and S_2 are no longer independent of concentration. In that case S_1 and S_2 will be indicated in what follows by the notations S_1' and S_2' and computed by means of equation (5). Discussing our measurements we could, however, observe that $[S - S_1' \cdot x^2 - S_2' \cdot (1-x)^2]/x \cdot (1-x)$ does not depend on concentration, which proves that the expression

$$\frac{I}{\lambda} \cdot \frac{d\beta}{dI} + \frac{I^2}{2\lambda(\lambda+1)} \cdot \frac{d^2\beta}{dT^2}$$

is small. * Thus we can put $\alpha = \beta$ approximately, and obtain in this manner some indication of the interaction between the components at low temperatures.

---

* Admitting for β the following expression: $\beta = a + \dfrac{b}{T^2} + \dfrac{c}{T^4}$, we can observe that

$$\frac{I}{\lambda} \cdot \frac{d\beta}{dT} + \frac{T^2}{2\lambda(\lambda+1)} \cdot \frac{d^2\beta}{dT^2}$$

becomes small when λ is approximately equal to 3/2.

We proceed in the following way:—We measure the velocity of sound in the gas mixture as a function of pressure at different fixed temperatures. From these measurements, we compute by the method of least squares the different values of S corresponding to the different concentrations. S_1 and S_2 are in most cases known from previous measurements carried out on the components, or $S_1(S_1')$ and $S_2(S_2')$ are computed from the temperature-variation of B given by measurements on isotherms.

The gas mixtures are prepared within a large vessel of capacity 4 or 10 l. respectively for the measurements at liquid-oxygen or liquid-hydrogen temperatures. The concentration is at first determined by means of the partial pressures. The mixture remains in this vessel for several hours. After this the
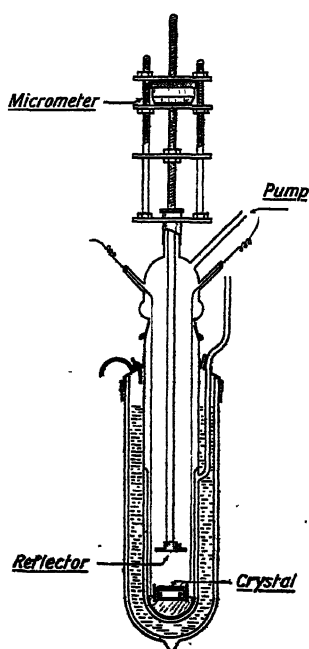
Figure 1. Acoustic interferometer.

gas is let into the acoustic interferometer represented in figure 1. The velocity of sound is measured in the mixture at room temperature and from these measurements the concentration is finally computed from equation (4). At low temperatures the concentration is checked again by using the value of W_0 obtained from the measurements represented by equation (3). The values of x found at room temperature and low temperatures agreed generally very well.

For the measurements on the pressure dependency of the velocity of sound, the pressure is lowered in the interferometer by letting the gas stream into an evacuated vessel of capacity 5 l. The gas pressure within the interferometer is read on a closed mercury manometer by means of a cathetometer.

The frequency of the piezo-quartz used is 523·78 kc./sec. For the details of the electrical system, we refer to our previous publications.

§3. MEASUREMENTS ON THE GAS MIXTURES A–He

The values obtained for the velocity of sound as a function of pressure, temperature and concentration are given graphically in figure 2.
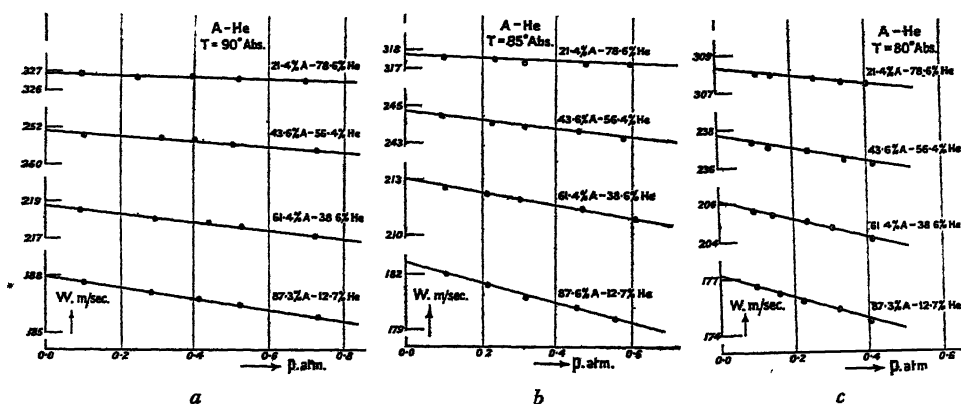
Figure 2.   Velocity of sound in A–He as function of pressure.

Using these experimental data we have computed the factor S and also W_0 corresponding to equation (3). $W_{0\text{ theor.}}$ is the value of W_0 computed by means of equation (4), x being the concentration observed at room temperature. It will be seen that the agreement between experimental values and theoretical ones is very good.

Table 1

T ($^\circ$K.)	Concentration (%)		$W_{0\text{ exp.}}$	$10^3 \cdot S$	$W_{0\text{ theor.}}$
	A	He			
90	87·3	12·7	187·9	−5·24	187·8
	61·4	38·6	218·7	−3·25	218·8
	43·6	56·4	251·8	−1·85	251·8
	21·4	78·6	326·9	−0·53	326·7
85	87·3	12·7	182·4	−8·44	182·5
	61·4	38·6	212·9	−4·64	212·6
	43·6	56·4	244·6	−2·75	244·7
	21·4	78·6	317·7	−0·84	317·5
80	87·3	12·7	177·2	−0·28	177·1
	61·4	38·6	206·1	−6·49	206·3
	43·6	56·4	237·6	−4·01	237·4
	21·4	78·6	308·2	−1·69	308·0

The values S_1–A and S_2–He are taken respectively from the measurements of van Itterbeek and van Paemel (1938) on the velocity of sound in pure argon and from the measurements of Nijhoff (1928) on the isotherms of helium gas. These values are indicated in table 2.

The values of S of table 1 are drawn graphically in figure 3, the full line corresponding with the theoretical curves computed by means of the equation

$$S = S_1 \cdot x^2 + \beta x \cdot (1-x) + S_2 \cdot (1-x)^2.$$

Table 2

T ($^\circ$ K.)	$S_1 \cdot 10^3$ A	$S_2 \cdot 10^3$ He	$\beta \cdot 10^3$
90	$-6 \cdot 60$	$0 \cdot 50$	$-3 \cdot 4$
85	$-10 \cdot 3$	$0 \cdot 38$	$-4 \cdot 3$
80	$-12 \cdot 6$	$0 \cdot 28$	$-7 \cdot 0$
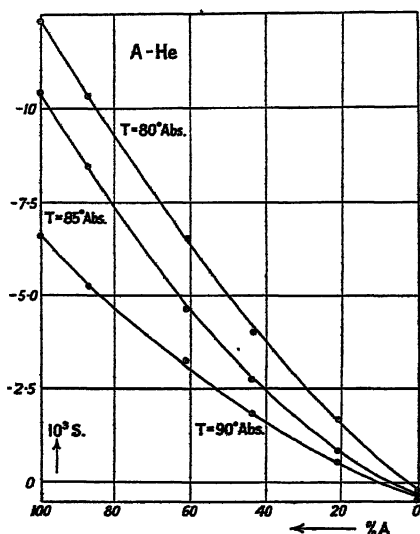
Figure 3. S in mixtures of A and He.
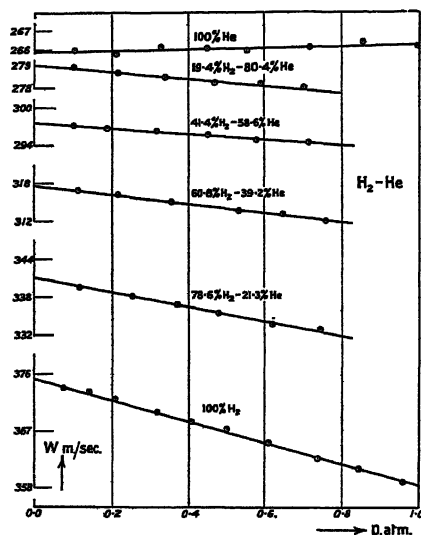
Figure 4. Velocity of sound as function of pressure at 20° K.

The interaction coefficient β is indicated in the third column of table 2.

§4. MEASUREMENTS ON THE GAS MIXTURES He–H$_2$ AT THE BOILING POINT OF LIQUID HYDROGEN

The values obtained for the velocity of sound as a function of pressure at 20°·0 K. are indicated in figure 4.

In table 3 (figure 5) we give the values of S and W_0 computed by means of the experimental values of figure 4. We see again that there is a good agreement between the values of $W_{0 \text{ theor.}}$ and $W_{0 \text{exp.}}$.

Table 3

T ($^\circ$ K.)	Concentration (%) H$_2$	He	$W_{0 \text{ exp.}}$	$10^3 \cdot S$	$W_{0 \text{ theor.}}$
20·3	0·0	100	375·5	$-3·39$	375·8
	78·6	21·4	341·1	$-2·44$	341·0
	60·8	39·2	317·7	$-1·65$	317·9
	41·4	58·6	297·6	$-1·06$	297·9
	19·4	80·4	279·2	$-0·43$	279·5
	100	0·	265·9	$+0·14$	265·4

We found again that S as a function of the concentration satisfies equation (8), with $\beta = -2\cdot11\cdot10^{-3}$.
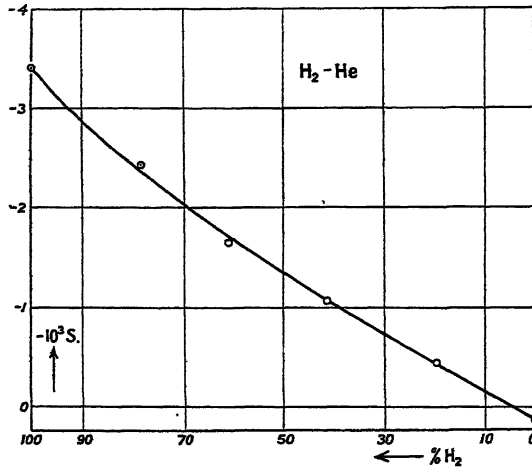
Figure 5.   Comparison of theoretical and observed velocity of sound.

§ 5. MEASUREMENTS ON A–H$_2$ MIXTURES

The experimental results for the A–H$_2$ mixtures are given graphically in figure 6.

The values of S, $W_{0\,\text{exp.}}$ and $W_{0\,\text{theor.}}$ computed by means of the experimental data are given in table 4.
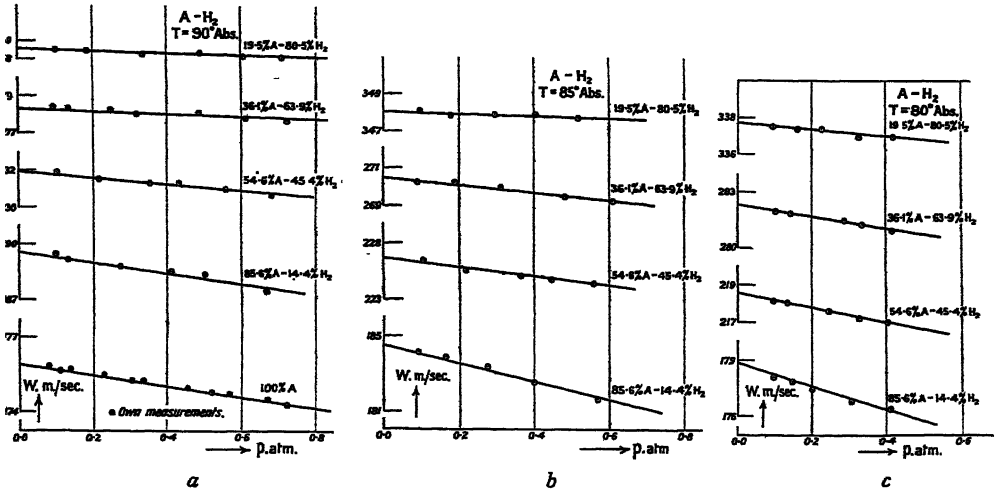
Figure 6.   Velocity of sound in A–H$_2$ mixtures.

The values of S_2' are computed by means of the formula

$$10^3 \cdot B = 0\cdot8742 - \frac{55\cdot818}{T} - \frac{5\cdot1515 \times 10^3}{T^2} + \frac{1\cdot0099 \times 10^5}{T^3} \qquad \ldots\ldots(10)$$

published by Keesom and van Lammeren (1932).

Table 4

T (° K.)	Concentration (%)		W_0 exp. (m./sec.)	$10^3 . S$	W_0 theor. (m./sec.)
	A	H_2			
90	100·0	0	—	−6·6	—
	85·6	14·4	189·7	−5·21	189·7
	54·6	45·4	232·0	−2·38	232·0
	36·1	63·9	278·5	−1·36	278·2
	19·5	80·5	358·5	−0·36	385·2
	0	100·0	—	+0·29	—
85	100·0	0	—	−10·3	—
	85·6	14·4	184·5	− 8·02	184·3
	54·6	45·4	225·2	− 3·62	224·4
	36·1	63·9	270·4	− 1·98	270·2
	19·5	80·5	348·0	− 0·62	348·2
	0	100·0·	—	+ 0·24	—
80	100·0	0	—	−12·6	—
	85·6	14·4	178·8	− 9·85	178·9
	54·6	45·4	218·6	− 5·17	218·8
	36·1	63·9	262·3	− 3·36	262·2
	19·5	80·5	337·6	− 1·47	337·8
	0	100·0	—	+ 0·17	—
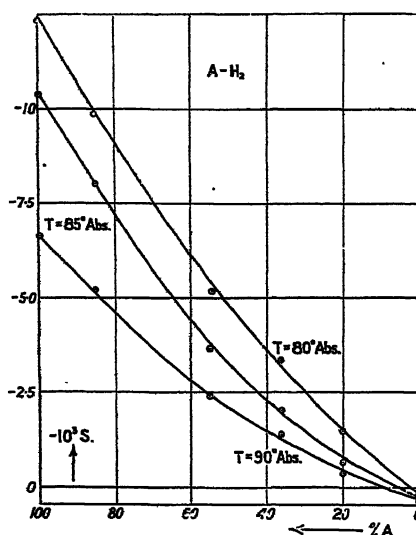
Figure 7.   Velocity of sound in A–H_2 mixtures as functions of temperature.

We computed β by means of equation (8), replacing S_1, S_2 by S_1' and S_2'. The values of S_1' are calculated by means of the equation

$$10^3 . B_A = 0·129 + 0·5923 \frac{10^3}{T} - 3·5923 \frac{10^5}{T^2} + 3·7269 \frac{10^7}{T^3} - 1·40030 \frac{10^9}{T^4} \quad \ldots\ldots (11)$$

published by van Itterbeek and van Paemel (1938). The values of S_2' are

computed by formula (10), the specific heats for hydrogen being taken from the book of Farkas (1935).

The values obtained for β for the different concentrations are given in table 5.

Table 5

T ($^{\circ}$ K.)	x (%)	$10^3 . \beta$	$10^3 . \beta$ average
90	80	−1·7	
	70	−1·8	
	60	−1·7	
	50	−1·8	−1·8
	40	−1·8	
	30	−1·8	
85	80	−4·8	
	70	−4·7	
	60	−4·5	
	50	−4·3	−4·5
	40	−4·3	
	30	−4·2	
80	80	−8·5	
	70	−8·2	
	60	−7·7	
	50	−8·3	−8·4
	40	−8·4	
	30	−7·9	

We see, as remarked in § 2, that β does not depend upon pressure within the limits of experimental accuracy.

§ 6. MEASUREMENTS ON A–D$_2$ MIXTURES

In order to compare the interaction of H$_2$ with that of its isotope D$_2$, we have also made some measurements on A–D$_2$ mixtures. These measurements, indicated in figure 8, are only made for a concentration at which the influence of the interaction is largest.
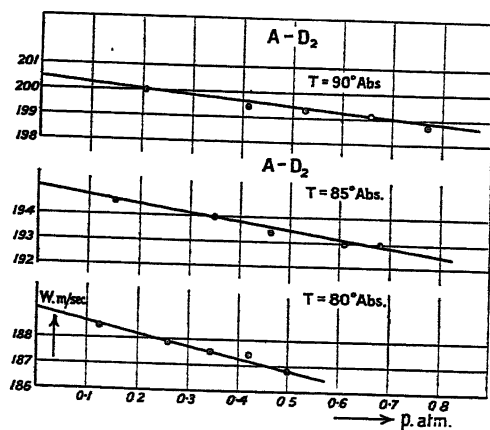
Figure 8. Results for A–D$_2$ mixtures.

From the experimental data of figure 8, we have computed the values of S and $W_{0 \text{ exp.}}$ (see table 6). $W_{0 \text{ theor.}}$ was calculated by means of the values of the specific heats of D_2 taken from Farkas.

Table 6

T (° K.)	Concentration (%)		$W_{0 \text{ exp.}}$	$10^3 . S$	$W_{0 \text{ theor.}}$
	A	D_2			
90	68·4	31·6	200·5	−3·64	200·7
85	68·4	31·6	195·1	−5·34	195·0
80	68·4	31·6	189·2	−7·34	189·3

In table 7 we compared the values of S obtained for A–D_2 mixtures with those of A–H_2 mixtures. It will be seen that there is no difference within the limits of experimental error.

Table 7

T (° K.)	$10^3 . S$ A–D_2	$10^3 . S$ A–H_2
90	−3·64	−3·55
85	−5·44	−5·45
80	−7·34	−7·40

Table 8

Mixture	T (° K.)	$10^3 . B_1$	$10^3 . B_2$	10^3
A–H_2	90	− 8·0	−0·2	−1
	85	− 9·0	−0·3	−4
	80	−10·0	−0·6	−8
A–He	90	− 8·0	0·5	−3
	85	− 9·0	0·38	−4
	80	−10·0	0·28	−7
H_2–He	20	− 6·56	0·10	−2

§7. CONCLUSIONS

In table 8 we have summarized our results obtained for the interaction term and compared them with the interaction between the individual components Thus we can observe that β is generally of the same order as the B-value of the component with the higher boiling point. Furthermore, when the temperature decreases, β increases very strongly.

Further work is in hand on H_2–N_2, H_2–He, H_2–O_2 and He–O_2 mixtures.

ACKNOWLEDGMENT

REFERENCES

FARKAS, A., 1935. *Light and Heavy Hydrogen* (Cambridge).

KEESOM, W. H. and VAN LAMMEREN, J. A., 1932. *Comm. Leiden*, 221 c.

NIJHOFF, G. P., 1928 *Comm. Leiden*, Suppl., 64 c.

VAN ITTERBEEK, A. and KEESOM, W. H., 1930. *Comm. Leiden*, 209 a.

VAN ITTERBEEK, A. and VAN DONINCK, W., 1943. *Physica*, 10, 481 ; 1944. *Ann. Phys* · (*Chim.*), 19, 88.

VAN ITTERBEEK, A. and VAN PAEMEL, O., 1938. *Physica*, 5, 593, and 848.

REVIEWS OF BOOKS

Optique Physiologique : La Dioptrique de L'Œil et sa Correction, by Y. LE GRAND.
Pp. 356 and 105 diagrams. (Paris: Revue d'Optique, 1946.)

The book under review represents Part I of a treatise by M. Le Grand on Physiological Optics. He is here dealing more particularly with the optical system of the eye on Gaussian principles, and with the method of testing and correction of refractive errors. The main contents of the book will be generally familiar to those who have studied visual optics, but the material is presented in a very clear and orderly manner. A necessary amount of algebra is included, but the text is not overburdened with it, neither is the clarity of the diagrams obscured by too many rays of light or too many details of the optical systems.

Since M. Le Grand has for many years been engaged in teaching physiological optics, it is not surprising to find a list of exercises given at the end of the book. As the solutions of the problems, and not merely the answers, are also included, it is possible for the conscientious student both to test and teach himself by this means.

A more unusual, but none the less welcome, feature in a book of this type is the inclusion of two appendices dealing respectively with elementary statistics and with the calculation of matrices in optical work. Whether the average optician will consider that he has any need to venture into these fields is perhaps a matter of doubt, but at least the material is provided in a readily assimilable form for those who have the inclination to attack it.

It is interesting to note that in the list of original papers quoted in the bibliography at the end of the book, half of them appeared in the *Transactions of the Optical Society*—a small matter, perhaps, but significant that the Society in its time performed a useful function and might have continued to do so if sufficient support had been forthcoming to keep the Society alive. Readers in this country will also be interested to see that Fincham's theory of accommodation is described as one of the most satisfactory.

Quite obviously M. Le Grand regards his subject as one worthy of very serious study by optical designers. As he remarks on p. 345, the modern correcting lens is a complicated optical instrument and the last word has not yet been said about the best design of spectacle lenses. M. Le Grand can at least claim to have given a clear account of the basic principles which must govern any future developments in such lens design.

W. D. WRIGHT.

A Survey of General and Applied Rheology, by G. W. SCOTT BLAIR. Pp. 196.
(London: Sir Isaac Pitman and Sons, Ltd., 1946.) 18s. 6d.

Rheology is defined as the study of the deformation and flow of matter. Thus defined it should obviously cover a large proportion of the field of physics. In practice the subject is concerned almost exclusively with those types of matter which do not obey the simple laws of deformation or flow to be found in the elementary textbooks of physics.

Dr. Scott Blair's book is written primarily for the scientist in industry, who has to deal with matter as it is rather than with matter as the physicist would like it to be. The first part of the book gives an account of the wide variety of rheological phenomena encountered in industrially important materials, and discusses the various types of apparatus by which these phenomena may be investigated. Those wishing to pursue the subject in greater detail will find the annotated bibliography particularly helpful.

The second, and more controversial, part of the book deals with the interpretation of rheological data. The central question here discussed may be stated in the following way: What physical factors are responsible for the formation of a subjective judgment of a given rheological property on the basis of sensory impressions ? Dr. Scott Blair has made extensive investigations into this question, and his discussion of this and related problems makes interesting reading.

L. R. G. TRELOAR.

RUTHERFORD : LIFE AND WORK AFTER THE YEAR 1919, WITH PERSONAL REMINISCENCES OF THE CAMBRIDGE PERIOD

BY J. D. COCKCROFT, F.R.S.

Second Rutherford Memorial Lecture, delivered 8 February 1946

§ 1. INTRODUCTION

THE Council of the Physical Society in establishing the Rutherford Memorial lecture directed that the first two of these should be reminiscent and personal. This would, I feel, have met with approval from Rutherford, whose love of personal reminiscences, and particularly of indiscretions, was a well known characteristic; so I shall begin by saying what qualifications I have for speaking of Rutherford and his laboratory in this personal way.

I first met Rutherford in 1914 in Manchester, where I had gone to sit at the feet of that most charming of mathematicians, Horace Lamb. As some light relief I attended the first-year lectures in physics. These lectures increased steadily in noisiness until one day the storm broke, and Rutherford was brought in to restore order. I still remember the immediate impression that here was a great man who was not going to stand any nonsense; thereafter the lectures were delivered by Rutherford in perfect quiet except for the applause which greeted the beautiful demonstrations of Kay, the laboratory steward. This was the first year of war, and my contact with Rutherford was broken for about seven years. I was then persuaded by Miles Walker in 1922 to take up again the study of mathematics in Cambridge, and was given the best of introductions to Cambridge, a letter to Rutherford. I remember going to see him in the old Maxwell Wing of the laboratory and finding him sitting, as he so often did, on a stool. He received me very kindly, and gave me authority to devote such time as I could spare from mathematics to work in the advanced practical class where, under Thirkill and Appleton, I learnt the elements of laboratory arts. After two years of the Tripos I was accepted as a member of the laboratory, and set to work in Chadwick's introductory training course. I worked first with Joseph Boyce on building and using the favourite tool of the Cavendish, the gold-leaf electroscope. After that we blew and built with our own hands a fine Macleod gauge—and were horrified by destroying it in the classical manner just as Rutherford walked in at the door. He was sympathetic, but explained that he did not keep students who had too many accidents. This dislike of sins against apparatus was one of Rutherford's well marked characteristics.

By this time, in 1924, Rutherford's Cambridge Research School was well established. The story of his second coming to Cambridge has been told in Eve's Biography: after the first approach from the University there was a typical Rutherford letter to J. J. Thomson asking for a clear mutual understanding with regard to the laboratory and research students. These negotiations went on for some time, helped by Sir Joseph Larmor acting as intermediary, and finally the conclusions were embodied in a document which formed the working agreement between Thomson and Rutherford for their respective spheres of influence in the Cavendish Laboratory. This agreement is reproduced in figure 1, whilst figure 2 shows the two signatories in later years. As a result of this agreement "J.J." retained the ground floor of the wing of the Cavendish built by Lord Rayleigh, a space always referred to as "The Garage". There he worked with his personal assistant Everett, who taught us glass blowing and sold to us boxes of "Everett wax". Every day at 1 o'clock or at the end of the day we would see J. J. standing at his little desk looking over the results of Everett's experimenting, or peering through his spectacles at the maze of glass tubing, which formed his experiment. He was a much-loved and venerable figure, always giving us a cheerful grin on passing, and always full of talk on topics other than physics.

Within the J. J. domain worked Aston, his research on isotopes then coming into full flower. During the post-war years he was able to separate the isotopes of neon, chlorine, mercury and lithium and to build thereafter the first of the precision spectrographs, shown in figure 3.

§ 2. BEFORE 1930

Rutherford's interests during the first decade of his work at Cambridge followed three main lines. First came his work on the transmutation of the lighter elements. This work was a continuation of that started in Manchester during the war years. An observation of Marsden in 1914 had shown that long-range particles were produced when alpha particles were fired into hydrogen gas. Working with Kay, the Manchester Steward, Rutherford showed that these particles were hydrogen nuclei. He went further and showed that similar particles were produced when alpha particles were fired into nitrogen. He concluded that a transmutation of nitrogen had been achieved. These experiments, published in the *Philosophical Magazine* in 1919, were a fitting climax to Rutherford's work in Manchester.

The work was taken up at once at Cambridge in Maxwell's old research room. Figure 4 shows the apparatus which he used with Chadwick. The alpha particles from the radioactive source passed through a brass chamber filled with nitrogen or other gases. The protons ejected were detected by the simplest of devices, a zinc sulphide screen. Between the screen and the chamber, absorbing screens of different thicknesses could be inserted.

Figure 5 shows Rutherford with his apparatus. Every day round about 4 o'clock periods for counting were arranged and two of the research students with good eyes were brought in to help. They sat in the darkened room for half an hour, and over cups of tea listened to Rutherford talking about all things

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

under the sun. Thereafter when Crowe had produced the source they took turns of a minute each to look through the microscope and to count the faint proton scintillations.

Mr. G. R. Crowe, who was Rutherford's assistant, gave to Eve a lively description of the work done at that time. He said, " He was shooting protons out of light atoms by means of alpha particles; carbon would not play ". The next day was the turn of aluminium, and on some theoretical ground Rutherford thought the protons from aluminium would have a high velocity and a long range. " Now, Crowe, have some mica absorbers ready tomorrow with a stopping power of 50 cm."——" Yes, sir ". " Now, Crowe, put in a 50 cm. screen "—— " Yes, sir ". " Why don't you do what I tell you; put in a 50 cm. screen "—— " I have, sir ". " Put in 20 more "——" Yes, sir ". " Why the devil don't you do what I tell you; I said 20 more "——" I did, sir ". " There's some damned contamination; put in two 50's "——" Yes, sir ". " Ah, it's all right; that's stopped 'em. Crowe, my boy, you're always wrong until I've proved you right. Now we'll find their exact range ".

Ellis, in his Obituary of Rutherford,* has described how firmly Rutherford kept control of the experiments in the darkened room. As the scintillations were recorded, Rutherford would compare them with what he expected and translate them at once into the final deductions. During the years up to 1925, transmutation of the light elements up to potassium was studied, and the ranges of the disintegration protons determined; the identity of the disintegration particles was proved with characteristic thoroughness by measuring their deflection in a magnetic field.

During this period Blackett embarked on his courageous attempt to get a Wilson chamber photograph of the nitrogen disintegration. He took photographs of hundreds of thousands of tracks and was eventually rewarded (figure 6) by seeing the tracks of disintegration protons. The photograph shows a long proton track shooting downwards. An alpha particle has entered a nitrogen nucleus, ejecting a proton and leaving oxygen behind, producing the transmutations $^{14}_{7}\text{N} + ^{4}_{2}\text{He} \rightarrow ^{17}_{8}\text{O} + ^{1}_{1}\text{H}$.

This was the first direct proof that the alpha particle is captured in the transmutation.

The second main line of experimenting in the laboratory in this decade was a study of the fields of force round atomic nuclei. The first simple scattering experiments of Rutherford and Marsden in 1907 had led to Rutherford's fundamental concept of the nuclear atom. These experiments were continued in the post-war years by the more refined experiments of Chadwick and Bieler. They shot alpha particles of varying velocity through gases and foils, measured the angular distribution and compared it with that to be expected if the inverse square law held. They were able to show that near heavy nuclei, such as gold, the inverse square law held up to the limits of penetration of the fastest alpha particles available. But with the lighter elements they found very marked deviations. These experiments presented Rutherford with the paradox that the potential barrier round the heavy nuclei must rise well beyond 10 million volts, whereas alpha particles come out of these nuclei at energies between

* *Proc. Phys. Soc.*, **50**, 463 (1938).

4 and 8 million volts.   The alpha particles do not therefore escape with the full
energy of the potential field.

The solution to this problem was provided by a visitor to the laboratory;
a young and imaginative theoretical physicist from Leningrad, Dr. Gamow.
Gamow saw that the solution was provided by the new wave mechanics which
was then being developed by Schrödinger and others.   On the new ideas it
was evident that the alpha particles could leak through the potential barriers
of the nuclei, and so need not have the full energy of the field.   In this way,
nuclear physicists were given a new and most fruitful model to guide their
researches.

The third main line of the laboratory during this period was the continuation
of Rutherford's Manchester and McGill interests in the radiations from radioactive
substances.   Ellis applied the magnetic focusing method to the study of the
gamma rays from the radioactive substances.   He was able to show that the
gamma rays resulted from transitions between nuclear energy levels.

These gamma-ray observations were linked in the laboratory with a study
of the alpha-particle spectra.   We built for Rutherford a special annular magnet
whereby he could study the fine structure of these lines.   The nuclear energy
levels so determined agreed in general with those found by Ellis's gamma-ray
work.

(At this point, the lecturer had records played in which Rutherford himself
was heard speaking of this work.)

§ 3. AFTER 1930

The second decade of Rutherford's work in Cambridge opened with an
intensive development of new laboratory techniques.   In various remote corners
and cellars, Wynn-Williams and his co-workers had been developing methods
for the electrical recording of particles.   In other parts of the laboratory, Feather
and Nimmo, E. J. Williams, Terroux and others were building cloud chambers,
spreading the influence of C. T. R. Wilson throughout the laboratory.   Allibone
came in from Metropolitan-Vickers, bringing with him a fierce Tesla coil which
produced 500,000-volt sparks, to the annoyance of the Corpus dons across the
way, and proceeded to build in the laboratory a high-voltage tube for the accelera-
tion of electrons.

Rutherford's Royal Society Presidential address of 1927 had urged the
development of sources of atoms and electrons with an energy far transcending
that of the alpha and beta particles from radioactive matter.   Allibone's 300-kilo-
volt electrons were the first step on the way.

Walton came to us from Trinity College, Dublin, and at Rutherford's
suggestion started to work on a new method of accelerating electrons by spinning
them round in the electric field produced by a changing magnetic field.   This
was the principle of the betatron.   Walton was able to develop the theory of the
stability of the orbits which has since been used as the basis of design of success-
ful betatrons.   The experimental problem was attacked with very limited
resources and Walton was not able to push the work through to a successful
conclusion.
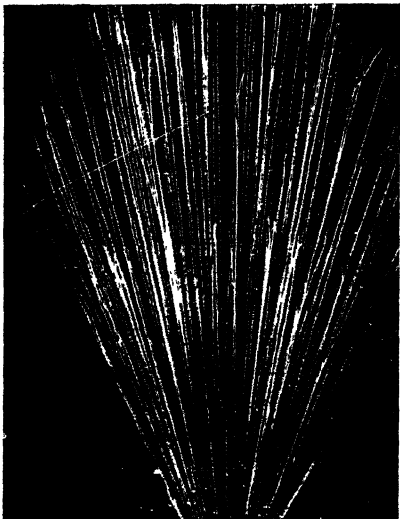
Figure 6.
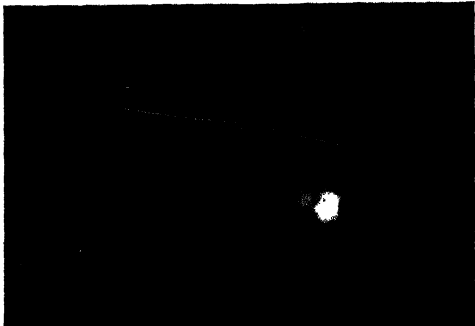
Figure 8.

Figure 7.

Figure 9.

Figure 10 (a).
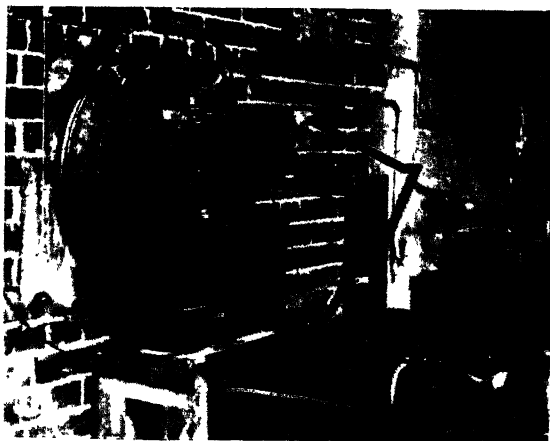
Figure 10 (b).

About the same time I started to build the first high-voltage apparatus for the acceleration of protons. I was led to do this by the predictions of Gamow's theory that protons of a few hundred kilovolts should penetrate the barriers of light nuclei. These predictions were submitted to Rutherford and he encouraged me to go ahead.

The fruit of all this development of technique came in that very wonderful year of 1932. First of all came the dramatic discovery of the neutron by Chadwick. For some time the laboratory had been interested in the radiations emitted when various light elements such as beryllium were bombarded with alpha particles. The radiations from beryllium were at first thought to be gamma rays, but did not behave in the expected manner, and for several months experiments were going on to try and find out what the radiation was. The announcement by Curie and Joliot of the projection of hydrogen nuclei by this radiation gave the clue to Chadwick. By a series of very rapid experiments in which the projection of different nuclei was studied, he was able to establish without doubt that the recoil protons were not due to a Compton effect, but resulted from the impact of material particles of unit mass and zero charge.

Very swiftly, Dee and Feather were set to work with their expansion chambers to look for the tracks of the projected particles. Dee found the tracks of proton and other recoils and Feather discovered the transmutation of nitrogen by neutrons, the first of the transmutations by this new particle.

Whilst this was going on, Walton and I had been completing our apparatus for the acceleration of protons. We had everything stuck together with plasticene by January 1932, and started to experiment with our proton beam, which passed out from the tube and travelled over 1 cm. in air. Figure 7 shows the apparatus. Rutherford was a little impatient at this stage, being very anxious to see the results. We put in a lithium target and a scintillation screen and were rewarded at once by seeing the bright scintillations characteristic of alpha particles. (In figure 7 Walton is seen sitting in the box with the recording apparatus used for looking for particles from disintegration.) We called in Rutherford at once and he was very enthusiastic. A few days later he announced the results with great gusto at a discussion of the Royal Society.

Rutherford's sound strategical sense was typified by his immediate diversion of P. I. Dee from his work with C. T. R. Wilson to join with Walton and myself in the study of the disintegrations by the cloud-chamber method. In figure 8 is a very beautiful picture taken by Dee in which the two alpha particles from one of these transmutations are seen travelling in opposite directions, whilst figure 9 is another Wilson chamber photograph showing the three alpha particles (joined by arrows) coming out of the transmutation of boron. The boron unites with a proton and splits up into three alpha particles.

Following on this, Oliphant was diverted from his work on positive ions to apply his technique to the production of strong proton beams for these experiments. Rutherford had to have an apparatus of his own to work with. Figure 10 shows the low-voltage apparatus built by Oliphant and Kempton in which 100 microamperes of protons were accelerated by about 200,000 volts.

About the same time in 1933 Rutherford received from G. N. Lewis the first sample of heavy hydrogen to reach Europe. Within a few days the sample was converted into deuterium gas, which Rutherford guarded with the most jealous care. He turned at once with Oliphant to do experiments with deuterons and soon discovered the transmutation of deuterium by deuterons. They found that two reactions occurred, $D + D = {}^1H + {}^3H$ or ${}^3He + {}_0^1n$, leading to the new elements ${}^3H$ and ${}^3He$. Two proton groups were discovered in the first reaction, and a short-range alpha-particle group in the second.

Rutherford and Oliphant's speculative interpretation of the reactions was confirmed by Dee's Wilson-chamber photography, Dee being lent for a few days a few cc.s of the precious gas. Figure 11 shows the tracks from the reaction ${}_1^2D + {}_1^2D = {}_1^1H + {}_1^3H$. The track on the left is a proton; the track on the right is the new kind of hydrogen of mass 3, a hydrogen which we now know to be radioactive, with a very long life.

One other discovery of major importance was made during this same period. During 1932 we had a visit from Millikan, who brought with him some very remarkable cosmic-ray photographs taken by Anderson which gave the first indications of positive electrons. Immediately after that, Blackett and Occhialini turned on their Wilson chamber to the search. Introducing the principle of counter control, they very soon obtained some remarkable photographs, showing pairs of positive and negative electrons, and in some cases showers of particles. Figure 12 shows one of these photographs.

This was followed up by Chadwick, Blackett and Occhialini's discovery of the production of electron pairs by high-energy gamma radiations, and finally, to close this sequence of exciting discoveries, there came Chadwick and Goldhaber's discovery of the photoelectric disintegration of deuterium by gamma rays.

The discovery of artificial radioactivity was missed in the laboratory, largely because we did not in general work with Geiger counters, and were looking for particles either with scintillation screens or with counters which would not respond to beta rays. When Curie-Joliot announced the discovery of the production of artificial radioactivity of alpha particle bombardment, Walton and I were able to borrow a Geiger counter equipment from Bainbridge and to show that protons could produce artificial radioactivity in carbon.

§4. THE DEVELOPMENT OF LOW-TEMPERATURE WORK IN CAMBRIDGE

The development of low-temperature work in the Cavendish Laboratory by Kapitza is one case where Rutherford backed strongly research work outside his own special interests.

Kapitza had come to England about 1922 on some Soviet Government mission. When I first met him in 1924 he was engaged in producing strong magnetic fields in small coils by discharging a home-made accumulator battery through the coil. The ground floor room of the block which is now devoted to Colloid Science resonated with the hammering of lead plates to straighten out defective
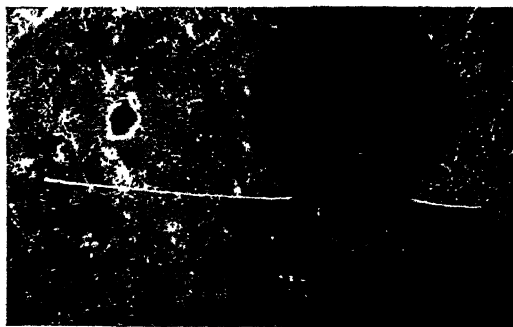
Figure 11.

Figure 13.
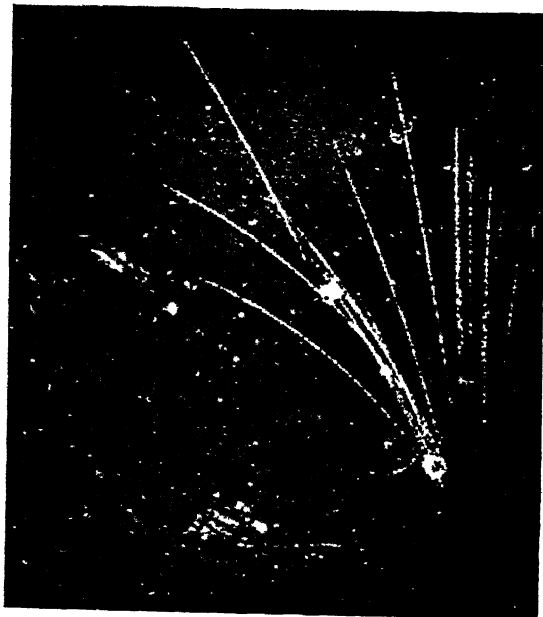
Figure 12. [With acknowledgments to Prof. P. M. S.
Blackett; reproduced by courtesy of Messrs. Long-
mans, Green and Co., Ltd.]

Figure 14.

Figure 15.
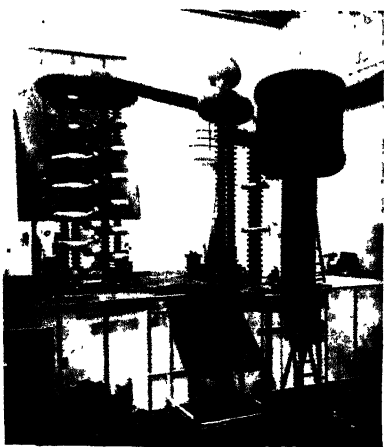
Figure 16.
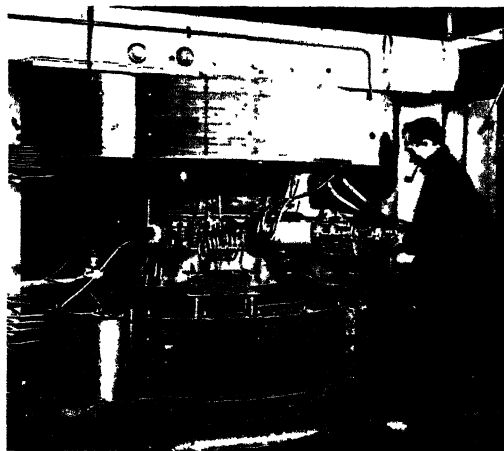
Figure 17.

members.   These fields had been used to curl alpha particle tracks and to show the loss of charge at the end.

Kapitza next designed an alternator to deliver currents of 20,000 ampères on short-circuiting the machine through his coils.   The D.S.I.R. provided the necessary funds, Metropolitan-Vickers designed and delivered the machine, and after many tribulations Kapitza built coils which stood up to the very high internal electromagnetic stress developed.

This led to the study of the electrical resistance of metals in strong magnetic fields.   The phenomena were found to be specially interesting at low temperatures, and so a hydrogen liquefier was built and installed in buildings wrung from the reluctant chemists.

After this, Kapitza determined to liquefy helium.   More space was required, and Rutherford was prevailed upon to seek the help of the Royal Society.   As a result the Society provided from the Mond Fund sufficient funds to build the Royal Society Mond Laboratory.   This laboratory was completed in 1933, and Kapitza built there the ingenious liquefier which for the first time eliminated the use of liquid hydrogen as an intermediary step in the liquefaction process.

Kapitza went to Russia in 1934 and did not return.   After some time Rutherford entered into negotiations with the Russian Government and Kapitza's equipment was transferred to the Institute for Physical Problems in Moscow which Kapitza built.   The Royal Society Mond Laboratory continued, however, to develop low-temperature work, and the discovery of new properties of liquid helium II was an exciting reward for the years spent in building up the new techniques in Cambridge.

§ 5. RUTHERFORD'S METHODS OF WORK

Finally I should say something about Rutherford's way of running the laboratory.   This was very characteristic of the man.   He devoted a great deal of thought to the selection of problems to be worked on by research students. Readers of Eve's *Life* will remember that in the Manchester days he used to carry in his pocket a long list of problems, and there were always plenty of spare problems.

In the later Cavendish period a list of problems was collected during the Long Vacation from the senior members of the laboratory, and Rutherford took this away with him to his country cottage and thought over it at leisure.   At the beginning of the October term the problems were ready to be assigned to students after their initial training period in Chadwick's attic course.

Throughout the year, however, Rutherford was always keenly alive to new situations.   I have already given you examples of the speed with which research workers were switched over to bring different techniques to bear on new problems.

With increasing duties, in London, Rutherford's tours of the laboratory become less frequent, but his visits and discussions, sitting as he used to do on a laboratory stool to go over the latest results, were always very welcome and

inspiring except for those rare days when he was in a black mood. These storms of Rutherford's used to blow up very suddenly, often started by mishandling of a favourite piece of apparatus. Then he raged through the laboratory and told everybody he met the worst that could be said about their habits of work and their experiments. Some of it was not fully justified, and after the storm it was all forgotten, but the prospect of these crushing analyses did quite a bit to keep the laboratory on its toes. Rutherford told me once he had mellowed a great deal in his later years. If so, these storms in Manchester must have been really awe-inspiring spectacles !

Figures 13 and 14 respectively show Rutherford sitting on a stool talking to Kempton and Wescott and to Ratcliffe on one of his tours of the laboratory.

One of the things that Rutherford never forgave was the publication of wrong results. For these he had a very long memory. He believed in the notice written in the entrance to McGill Physics Laboratory—" Prove all things ".

In his later years Rutherford was not able to engage with his own hands in experiments, but from time to time he would try his skill. I remember Crowe telling us about an occasion when he was trying to fix a thin mica window (which is apt to be a delicate operation). His hands were shaking rather too much to make it go. He said to Crowe: " Crowe, why the devil are you shaking the table ? "——" I'm not, sir ", said Crowe. " Crowe, why the devil are you shaking the table ? "——" I'm not, sir ". " Oh, well, Crowe, you come and fix it yourself ".

Rutherford did not believe in committees for running the laboratory. He had his own system which he called " polling the jury ". On all important matters of laboratory policy one or two senior members of the staff were summoned individually to his office and the matter was very thoroughly discussed. From the advice given to him Rutherford made up his own mind, and he was very rarely wrong.

The detailed work of the laboratory was divided up amongst the seniors. Thirkill did most of the supervision of accounts and represented the laboratory on the principal University Committees. Chadwick was responsible for the detailed supervision of the research students and for the provision of research apparatus. I looked after buildings and engineering equipment; Dee and Ratcliffe were in charge of Part II teaching and were responsible for the recommendations of students to go on to research. Very great care was devoted to the selection of these students. Experimental ability was always taken into account as well as paper results.

One of Rutherford's characteristics was his willingness to let his boys go ahead and develop new techniques and to spend quite a long time doing it— one or two years—without any appreciable interference or control. He took very little interest in the technical development; all he cared about was getting some useful results at the end of it. But as a result of this willingness to let people play, very marked contributions to the development of nuclear technique were made by Wynn-Williams, Lewis and others.

In a rather similar detached spirit Rutherford tolerated the other kinds of research work in the laboratory. He was never really interested in subjects

outside nuclear physics, but he recognized they had their place in the sun provided that place was small compared with that occupied by nuclear physics.

He was at his best in presiding over the fortnightly meeting of the Cavendish Physical Society. At the first meeting of the year he gave an enthusiastic review of the previous year's work of the Cavendish. Thereafter he would preside over the meetings, and however involved the presentation was, Rutherford would emerge, often from apparent slumber, with some pertinent question or with a masterly summing up.

Other characteristics of Rutherford's ways of life are well known. He was a great reader with wide tastes. On Sundays he was always to be found on the Gog Magog Hills, usually with Fowler, Aston and Taylor ; figure 15 shows this foursome in South Africa. On Sunday evenings he dined in Trinity, very often interchanging indiscretions with Fowler and others in the booming voice for which he was famous.

§6. THE NEW CAVENDISH LABORATORY

One of Rutherford's last duties in Cambridge was to prepare for the new building of the Cavendish. He took no particular pleasure in this since, as he often said, it was building for his successor. But he answered the request of the University to raise an endowment fund with characteristic ability. He called on Eddington to write his brilliant pamphlet on the Cavendish Laboratory ; he gave instructions for an official document to be prepared by myself; and he brought in the Chancellor to look for big money. Having obtained the great Austin benefaction, with characteristic promptness he sent his boys round the Continent at his expense to look at the best laboratories, and when they came back he gave them a pretty free hand in designing the laboratory. He took some pleasure in spending a little of the money on two high-voltage sets and on a cyclotron. " There won't ", he said, jingling his money, " be very much left for my successor ". Figure 16 shows the million-volt apparatus in the high-voltage laboratory, and figure 17 the cyclotron.

§7. CONCLUSION

Although Rutherford died in 1937, his influence was a major factor in the scientific supremacy of Britain in the war. The Senior Staff and research students of the laboratory, together with members of other physics schools, were mobilized in the first days of the war and developed for Britain and the allied cause the centimetric radar which turned the tide of the U-boat battle, directed the bombing of Germany and helped decisively to sink the Japanese fleet. The Liverpool branch of the Rutherford School, with Frisch and Peierls from Birmingham, initiated the work on the atomic bomb which ended the war with Japan. One can only wish that Rutherford had been alive to deal with its consequences.

THE EXPERIMENTAL BASIS OF ELECTRO-MAGNETISM: THE DIRECT-CURRENT CIRCUIT

By NORMAN R. CAMPBELL and L. HARTSHORN

ABSTRACT. The purpose of the enquiry, of which this paper forms the first part, is to show to what extent the working principles of electromagnetism can be soundly based on real experimental facts—that is to say, on experiments that have actually been performed, as distinct from the imaginary experiments, which are common to most expositions of the subject, but which are either quite impracticable or incapable of being performed with an accuracy that would be regarded as significant to-day. By maintaining this sharp distinction between fact and theory we aim at removing from the subject much of the confusion which, as the literature shows, is a continual source of trouble.

This section begins with an outline of the general principles of measurement. Current, resistance, conductance and voltage are then established independently of a knowledge of any other magnitudes. Ohm's Law and the conception of e.m.f. follow. An examination of the relations between these magnitudes and the geometrical and mechanical magnitudes then leads to theoretical conceptions like potential, and to a consideration of the status of Ampère's two laws. That for the force between two circuits is well established and defines the unit of current. That for the torque on a compass needle near a circuit is less well established ; but it leads to the recognition of H in a non-magnetic medium as a defined magnitude, whose significance is found to be independent of the existence of magnets.

§ 1. INTRODUCTORY

A REMARKABLE feature of the literature of the last decade is the number of papers that have been published on the fundamental magnitudes of classical physics. Even more remarkable is the divergence of opinion expressed on such well worn themes as the precise significance of the magnitudes of everyday electromagnetism. The experimentalist must occasionally ask himself whether ideas that have been successfully applied in an ever-widening field for so many years can really rest on such uncertain foundations.

There seems to be fairly general agreement that the working concepts are defined by the processes adopted for their measurement, but there is no sort of agreement as to what is the essential character of these processes. Some magnitudes appear to be associated with many processes, and there is no agreement as to their relative status.

Many difficulties have arisen from the fact that the basis of the subject has almost invariably been sought in processes that are not really practicable—that is to say, in theoretical experiments rather than in experimental facts. Experiments with point charges and magnet poles are well known examples. Every experimentalist soon learns that, whatever part such experiments may have played in the history of the subject, the best approximation he could make to

them would to-day be considered quite worthless as evidence in support of his working principles. He realizes also that disputes about such experiments are almost meaningless in relation to science as practised in the laboratory. The very fact that all the doubt and discussion about these theoretical ideas have had no discernible effect on the course of experimental enquiry is sufficient to show that there is no very close connection between the two. The world-wide agreement among physicists in practical matters shows that we must be working on some common basis. It is curious that it should be so difficult to express it in terms on which we are all agreed. The present authors are of the opinion that the difficulty can be overcome by considering those processes only that are actually employed in practice; and in this paper we propose to show that all the working laws of electromagnetism can be soundly based on experimental operations that will be generally recognized as forming part of present-day technique. Every physicist will agree that the basis of the subject is to be found in experimental facts. If, therefore, we can derive our working principles from experiments that are frequently performed in the normal course of physical enquiry, or that are at least representative of existing technique, we shall have established our basis clear of theoretical ambiguities.

Our enquiry shows that the basic experimental laws are not those which are given most prominence in the customary expositions of the subject. It is therefore the more desirable that their importance should be generally recognized, and particularly so in the work of establishing units and standards. The work of the various international conferences has been of great practical value in securing international uniformity, and it is of the utmost importance that it shall not become confused with irrelevant theoretical considerations. The international committees are primarily concerned with securing agreement on the basic facts: their decisions will only have their proper effect on scientific practice so long as the relation between these facts and our working principles is clearly understood.

§ 2. GENERAL PRINCIPLES

By saying that the laws are to be soundly based on the experiments we mean that the development must follow an order that may be described as logical, in the broad sense in which that term is normally used in scientific investigations: we must not assume the truth of any proposition until it has been established. Thus, we must not assume that we can establish any algebraic law relating different properties until we have shown that these properties are measurable by processes that do not depend on the truth of that law.

It follows that we cannot begin with any algebraic law. The subject necessarily starts with qualitative observations, which first become quantitative when we have devised an operation whereby we can make consistent judgments of equality in respect of some observed property, and another operation which possesses the characteristics of addition. We can then, by successive addition, build up a standard scale for that property, and make measurements by establishing judgments of equality between the things measured and determinate parts of the standard scale. The measurement of length is a very familiar example

of processes of this type, but it is not generally realized that the basic measurements of electrical properties are of the same kind.

Having established in this way the independent measurement of at least two properties, we are in a position to establish algebraic laws relating them—that is to say, relations between our observations of these two quantities that can be represented by algebraic equations. Such independent observations can never, except by the rarest fluke, show a relation of equality: they may, however, prove proportionality. It follows that the simplest algebraic law that we can establish will take the form $y = kx$, where x and y represent the two magnitudes that are independently measurable and k is a constant of proportionality which can be determined for any given system of the class to which the law applies from the measured values of x and y. We may now find that laws of the same form apply to other classes of similar systems, but that the constant k varies from one class to another, and is characteristic of some recognizable feature of the class. Thus k becomes established as a measure of this new property, which may be described as measurable by application of the law $y = kx$ to direct measurements of x and y. Such processes of derived measurement enable us to establish as magnitudes properties that are not additive, such as density and resistivity, and that may be described as *qualities*, to distinguish them from the additive properties or *quantities*. More complicated algebraic laws may take the form

$$y = kf(x_1, x_2, \ldots x_n),$$

where $f(x_1, x_2, \ldots x_n)$ denotes some definite function of any magnitudes $x_1 \ldots x_n$ that have already been established by either independent or derived measurement. We may regard $f(x_1, x_2 \ldots x_n)$ as a new magnitude z, which we shall call a *defined magnitude*. From an experimental point of view the law reduces to one of proportionality as before.

The value of the constant k in the above laws will obviously depend on the choice of units for x and y, as well as on the property which has been recognized as a derived magnitude. We may symbolize this fact by writing the equation in the form

$$y = S. a. f(x_1, x_2 \ldots x_n),$$

where a denotes the derived magnitude and S is the *scale factor*, which varies with the choice of units only. Sometimes the constant k turns out to be the same for every system to which the law is applicable. There is then no derived magnitude, for we always use the term " magnitude " to mean something that varies from one measurable system to another. In that case there is no need for our immediate purpose to split k into two factors; it can be regarded simply as a scale factor S; and if we decide to adjust our units of the xs and ys so that this S becomes unity, we can simplify the equation to the form $y = f(x)$. Even when k varies with the system, and a derived magnitude has to be recognized, we can always reduce our equation to this simple form for one particular class of systems by adjusting the units of the xs and ys so that $S. a$ becomes unity for this class. In other words, we adjust the units of the xs and ys until S becomes unity, and define the unit of a as the property possessed by that particular class. Examples will appear later.

We shall follow the customary practice of suppressing scale factors as far as we can consistently do so. If we have established m magnitudes, and n experimental relations between them, then by suitably choosing the units of the m magnitudes we can suppress m scale factors. If $n > m$, the remaining $(n - m)$ scale factors must appear as numerical factors in the equations. Whether derived magnitudes appear or not will depend only on whether such magnitudes have been established by the experiments.

However, it is sometimes desirable to introduce another symbol as well as the scale factor S, even when no derived magnitude has been experimentally established. For example, in a few instances the constancy of k has been established over a range of experiments so wide that it has acquired a special theoretical significance as a *universal constant*. It is well to guard against the suppression of such a constant by separating it from the scale factor. Again, we may have theoretical reasons for thinking that k may be variable, even though we have not succeeded in obtaining conclusive experimental evidence of significant variation. In such a case we may introduce a factor besides the scale factor so as to leave open the question whether k is actually variable. Obviously we are at liberty to introduce as many factors as we find to be significant in either theory or experiment, so long as we bear in mind the limitations to the significance of each.

There remains for consideration the order in which the various magnitudes should be introduced, and in this connection we note that our algebraic laws must always remain indefinite in their application, to an extent that is represented by the experimental errors of the observations on which they are based. Moreover, it is a well known fact that when the measurement of any property depends on the measurements of others, the experimental error of the dependent measurement is always greater than that of any of the ones on which it depends. A process of measurement that has proved in practice to be significant with high precision cannot therefore be soundly based on a law that has only been established with inferior accuracy. For example, it is absurd to suppose that dielectric constant, a magnitude that has proved to be significant in a very large number of physicochemical investigations with a precision better than 0.1%, is based on a law for the forces between small charged bodies which has never been established with anything like this accuracy; and it is even more absurd to suppose that voltage, a magnitude that is measured as a matter of routine by electric power companies with an accuracy of 0.01%, and in standardizing laboratories with a precision of 1 part per million, is to be defined in terms of the work done in moving an electric charge from one point to another, an operation about which our knowledge is incomparably more vague. We must therefore recognize that the familiar definitions, common to most expositions of electromagnetism, are the premises from which the mathematical reasoning proceeds, rather than definitions of the magnitudes with which we are concerned in experimental science; and in attempting to build up the subject from its experimental foundations we must begin with the independent measurement of the magnitudes that have been established with the least uncertainty, and then pass on to derived measurements based on them, always giving priority, in any one branch of

the subject, to the experiments that are known to be capable of the highest accuracy.

This means that our first stage must include independent measurements of resistance; for there can be no doubt that measurements of resistance can be made with a higher accuracy, and over a greater range, than is attainable for any other property that is recognized as being distinctly electrical in character; and, moreover, the most precise measurements of resistance are in fact independent measurements. Qualitative observations of current are, however, a necessary preliminary; and as the independent measurement of current and voltage is on the same footing as that of resistance, it will be convenient to regard these three magnitudes together as comprising the first stage, and to begin with current and the idea of the electric circuit.

We shall take for granted the existence of all the instruments commonly employed in physical laboratories. From our standpoint they are merely common objects of the modern world, just as easily identified, and far more common, than the traditional amber, catskin, and lodestone, which, indeed, have proved, on closer investigation, to be no less complicated in their structure; and, what is more to our present purpose, far less amenable to precise observation.

§3. CURRENT: A MEASURABLE PROPERTY

Construct a circuit of a most general kind by connecting in series a battery, a variable resistor and, say, a filament lamp, a neon lamp, a water voltameter, and indicating instruments of all the common types—moving coil, moving magnet, thermal, electrostatic (with shunt resistor), and electro-dynamometer or current balance.

Varying the resistor, we observe that the indications of all these devices increase and decrease together; there is something characteristic of the whole collection, and not merely of each element of it. By connecting the instruments with different materials we discover the existence of conductors and insulators, and learn that the characteristic " something " is associated with a definite closed path or circuit that must be conducting throughout. Further observation reveals the following facts:—

To any definite indication of one instrument, there are corresponding definite indications of the others, and these corresponding indications are independent of the order in which the various instruments are connected in the circuit, provided that each instrument is always connected in the circuit in the same direction with reference to the rest of the circuit. The reversal of some instruments gives rise to a change in their indications: others are unaffected by reversal. We shall describe instruments of the first class as *direction-sensitive,* and those of the second as *reversible.*

These observations suggest an analogy with the flow of water, the battery being analogous to a pump, the resistor to a throttle, the closed conducting path to a pipe, and the indicators to rate-of-flow meters of various kinds. The analogy further suggests that if we construct more complex circuits, in which several branches lead into a main conductor, the characteristic feature " current " of the main circuit may be the sum of the corresponding features of the branches:

in other words it suggests that there is a property "current", which is characteristic of electric circuits, and obeys a law of addition in branched circuits. We now establish this law and thereby establish current as a measurable property or magnitude.

We construct two circuits X and Y (figure 1) with a common limb, in which we connect our indicating instruments AA. For the purpose of simplifying the description it will be convenient to imagine that all the instruments possess pointers moving over scales of the ordinary type, which are, however, initially blank, but it will be readily understood that what we say about numbering the scales is equally applicable to scales of any type, such as a set of " weights " for the current balance, which we can mark with the appropriate numbers in due course, and so on. In the branches B and U we have also indicating instruments. U is the instrument we choose to " define our unit " : in other words, we assign the numeral 1 to some definite condition of this instrument. For this country
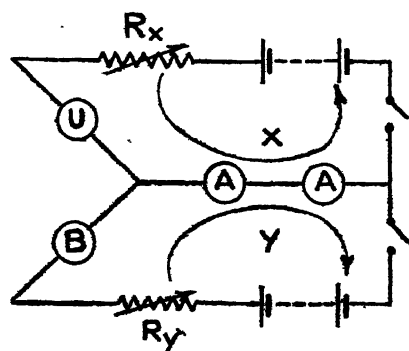
Figure 1.

the legal standard is a current balance, and the definite condition is one of equilibrium with a certain standard weight on one pan of the balance. B, like AA, has a blank scale. We first break the circuit Y, adjust R_x until U is in the standard condition, and then mark all the scales of AA with the numeral 1 at the points corresponding to their indications. We then open X and close Y, adjust R_y until AA again indicate 1, and mark the corresponding indication of B also 1. Next close both X and Y, adjust R_x and R_y until both U and B indicate 1, and mark the corresponding indication of AA with the numeral 2. Proceeding in the same way, we use the circuit Y to mark the corresponding indication of B with the numeral 2; then with B at 2 and U in the standard condition to mark AA at 3, and so on. Having marked all the instruments AA in this way, we can insert some in the place of U and others at B, and verify that whatever the indication of the three groups of instruments they always satisfy the law of addition. Further applications of the same principle enable us to subdivide the scales to any desired extent within the limits set by the sensitivity of the instruments, and then to verify the law with high precision for a very wide range of current. We have then established the indication of such an instrument as a unique and precise numerical measure of the current in any branch of a circuit in which it is included.

It is important to notice that the measurement of current in this way is completely independent of the law of operation of the indicating devices. We may notice that many direction-sensitive instruments have scales that lead to the recognition of some " linear law "; and that the scales of many reversible instruments suggest a " square law "; but our ideas of the magnitude " current " are not founded on any such law. The highest precision will be obtained when the indicating devices consist of the combination, well known in all standardizing laboratories, of a shunt resistor and a potentiometer, the indication being the setting of the potentiometer " slider " which corresponds with a state of balance of the potentiometer, but we must suppose that the scale traversed by the slider is initially blank, and that the combination is calibrated as an ammeter by the process we have described.

§ 4. CURRENT: A DIRECTED MAGNITUDE

The fact that the pointers of some instruments move in the opposite direction when the connections to their terminals are interchanged does not itself introduce any question of negative current; in effect we obtain a different instrument by the reversal, and must re-calibrate it, but that is all. Each direction-sensitive instrument must therefore have its terminals marked, say, red and black, and must always be connected in circuit in the same way with reference to the battery, so that its indications always correspond with those obtained during the calibration. If, now, bearing in mind this condition, we take the circuit of figure 1, and reverse both the battery in the circuit Y and the instrument B, we find, on closing both circuits X and Y, and adjusting R_X and R_Y until both B and U indicate the same current, that A then indicates zero current, whatever may be the current indicated by B and U. If our law of addition is to be generally applicable to branched circuits, we must therefore recognize that the sum of two equal currents may be zero; in other words, that there are such things as negative currents.

We adopt a simple convention to enable us to allocate signs to currents. We use direction-sensitive instruments, and mark their terminals so that during calibration the red terminal of one instrument is always connected to the black one of the next in the same circuit. Then at any junction the currents indicated by instruments with their red terminals connected to the common point are given the + sign, and those of instruments with the black terminals connected to the common point the − sign.

We now establish as an experimental fact Kirchhoff's First Law. The sum of the currents in all the branches meeting at any junction is always zero. This is one of our basic laws—grounded directly on observed facts and independent of any theory of electricity.

§ 5. RESISTANCE

The idea of resistance as a magnitude characteristic of the various parts of an electric circuit first arises when we observe that the insertion of a resistor in, say, the simple circuit first considered diminishes the indications of all the instruments, and that the insertion of a second in series with the first decreases it still further. We are led to define resistors as equal in resistance when the substitution of one for another causes no change in the indicator, and to establish the operation

ɔf connecting resistors in series as addition. It is scarcely necessary to enter into details. Every experimentalist knows that the most accurate measurements ɔf resistance are always made by " simple substitution ", which means by making the above-mentioned judgments of equality; and in standardizing institutions, the standard scale of resistance is always established by constructing resistance " build-ups "—that is to say, groups of resistors connected in series in such a way that it can be shown by experiment that they satisfy the law of addition with high precision. Resistance, then, is a property defined by these operations, and measured independently of any other property; for it may be noted that only qualitative observations of the current indicators are necessary. The various bridges and balancing devices employed for the work are to be regarded as detectors rather than measuring instruments—that is to say, they enable us to make precise judgments of equality of resistance much as a comparator enables us to make precise judgment of equality of lengths. Over a wide range of conditions, resistance proves to be independent of both the circuit employed for its measurement and the property *current*.

§ 6. CONDUCTANCE

It will be convenient to consider also· at this point *conductance*, another magnitude independently measurable, but closely associated with resistance; for the test of equality for the two quantities is exactly the same. They differ only in the law of addition, conductance being added by parallel connection instead of series. We find by experiment that for any resistor, conductance and resistance are inversely proportional to one another. The constant of proportionality is merely a scale factor which for convenience is suppressed by defining the unit of resistance and the unit of conductance by means of the same standard resistor. Conductance and resistance are then reciprocals.

§ 7. VOLTAGE

At first sight it would appear that we can establish a property " e.m.f." characteristic of batteries by a process almost identical with that established for resistance and equally fundamental, involving only addition by series connection and judgments of equality based on observations of no change of current on substitution in any circuit. Since the reversal of a battery produces changes of " current", we must distinguish the two terminals as, say, red and black, and when we find that the series connection of a battery in one direction increases " current ", and in the other direction decreases it, we are prepared to recognize e.m.f. as a directed magnitude, the reversal of the battery changing the sign. Detailed experiments show, however, that the method cannot be established. The judgments of equality are found to depend on the circuit employed as well as on the batteries under comparison; and with any one circuit, the combination (3–2) for example, is never found to be precisely equivalent ·to the combination (2–1). Subsequently we explain the failure of the method by recognizing another property of batteries, " internal resistance ", but at this stage we can only conclude that the fundamental property *voltage* must be sought elsewhere.

We find that we have two distinct classes of electrical indicating instruments. In addition to those that we have used in order to establish the magnitude *current*

(the ammeters), we have others which, unlike the ammeters, cannot form part of any circuit, since on inserting them into any circuit they invariably give rise to zero current, but which nevertheless give definite indications when connected across any component of a circuit.   We shall now show that just as the instruments of the first class enabled us to establish the magnitude current in virtue of a law of addition, those of the second class (the voltmeters) enable us to establish voltage by a similar law.   As in the first case, the process is quite independent of the law of operation of the instrument itself, but we may note that our voltmeters include electrostatic instruments, thermionic instruments of the electrometer type, and potentiometers.   Some of these contain batteries and are found to be direction-sensitive: others are reversible and obey a different law, but all serve to establish the same magnitude.   The potentiometer is by far the most accurate of the voltmeters.   The fact that the same mechanism also formed part of one of the ammeters need cause no confusion: we must again suppose that the instrument has initially a blank scale, which by appropriate marking and numbering will be established as a scale of voltage.

We observe that if any voltmeter is connected in succession to points XY, YZ, XZ (figure 2), Y being between X and Z and no battery being between X and Z,
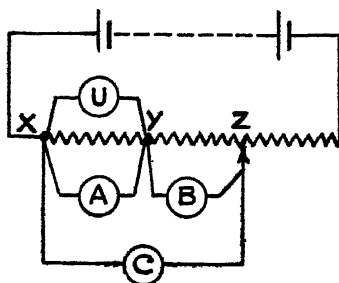
Figure 2.

then it gives a greater reading in the position XZ than in either of the positions XY or YZ.   This observation suggests that a law of addition can be established for the property indicated, voltage being additive for circuit elements connected in series.   The experimental procedure is closely analogous to the 3-instrument method used for current, and will be sufficiently obvious from figure 2.   One voltmeter U is arbitrarily chosen as standard, and its indication when connected to the points XY, say, is marked 1.   The indication of a second instrument A, also connected to XY, is also marked 1, and it is then transferred to B, and the point Z adjusted until it again reads 1.   At this setting the indication of the third instrument C is marked 2.   Proceeding in this way we establish our standard scale of voltage by successive additions of 1, and verify the laws of addition by proving that the indications of the three instruments always satisfy the relation $C = A + B$.

In order that the law of addition may apply perfectly generally to all circuit-elements in series connection we must recognize voltages of both signs.   Thus in the circuits of figures 3 and 4 a voltmeter connected to the terminals B and D may indicate zero, even when it gives finite indications in connection to BC and CD.

The sum of the voltages of BC and CD is by the law of addition zero, and, therefore, these two voltages must be equal but of opposite sign. Positive and negative voltages can be recognized by the use of direction-sensitive voltmeters with marked terminals, say one red and one black, as for ammeters. With such an instrument we find that connection to the terminals BC and CD in the case just mentioned gives the same indication provided that the same terminal of the voltmeter is connected to B of the pair BC and to D of the pair CD. Thus, reversal of the
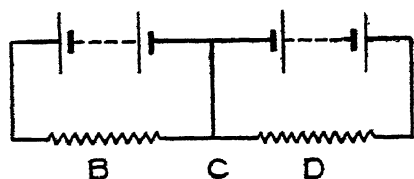
Figure 3.
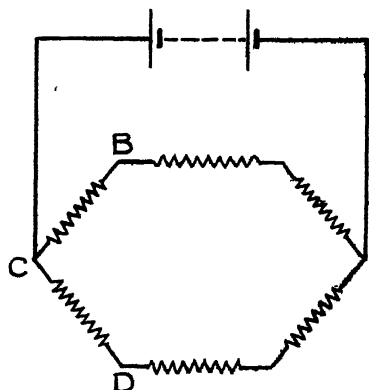
Figure 4.

order of the voltmeter terminals when passing along a circuit indicates a change of sign of voltage. We may note here that the terminals of any component of a circuit are essentially the devices by means of which a voltmeter can be connected to two definite points on that circuit-element.

We are now in a position to establish the following experimental law concerning voltage. The algebraic sum of the voltages measured between all the successive pairs of any set of terminals on any closed circuit is always zero. This law is analogous to Kirchhoff's First Law, and is of the same basic importance.

§8. OHM'S LAW

Having now completed our first stage and established the independent measurement of current, voltage and resistance by processes that will give, in favourable circumstances, an accuracy of the order of 1 part per million, we can now proceed to the development of algebraic laws based on these measurements. The most obvious one is Ohm's Law: for a large class of circuit-elements the terminal voltage V is proportional to the product of the resistance R and the current I characteristic of that element:

$$V = kIR. \qquad \ldots\ldots(1)$$

The proportionality constant k proves to be independent of the size and shape of the circuit element and of its composition; metals, non-metals, electrolytic solutions, etc., all have the same value if the law holds. The constant is therefore no more than a scale factor, and is suppressed by adjusting the units of V, I and R so that $k=1$. Incidentally the fixing of the sign of k determines the relation between the red and black terminals of ammeters and those of voltmeters, which have hitherto been unrelated.

§ 9. ELECTROMOTIVE FORCE

We now find that there is an important class of circuit elements, including batteries, dynamos, thermocouples, etc., which do not obey Ohm's Law, but are characterized by a finite terminal voltage even when they are " on open circuit " and, therefore, have zero current. When these devices are included in a circuit we find that their terminal voltage varies with the current, and we establish the experimental law

$$V_0 - V = kI,$$

where V is the voltage corresponding with the current I, and V_0 is the voltage on open circuit. The constant k proves to vary with the circuit element and to be characteristic of it. We must therefore recognize a new magnitude r as well as the scale factor S and write

$$V_0 - V = SrI \quad \text{or} \quad V = V_0 - SrI. \qquad \ldots\ldots(2)$$

If, then, the circuit external to the generator consists of n circuit elements in series, all obeying Ohm's Law and of resistance $R_1, R_2 \ldots R_n$, we have, by applying Ohm's Law and the law for the summation of the voltages of a closed circuit,

$$IR_1 + IR_2 + \ldots IR_n - (V_0 - SrI) = 0.$$

Note that a negative sign has been attached to the expression for the voltage of the generator in this equation. This is because experiment shows that the order of the voltmeter terminals must be reversed when passing from the resistor elements $R_1, R_2, \ldots R_n$ to the generator. When we examine this relation we perceive that a great simplification can be achieved by choosing the unit of r so that the scale factor S becomes unity, in which case the equation becomes

$$I\left[\sum_1^n R + r\right] = V_0. \qquad \ldots\ldots(3)$$

It is natural therefore to interpret r as a resistance inherent in the generator, and with this interpretation we obtain Kirchhoff's Second Law in the familiar form

$$\Sigma IR = E, \qquad \ldots\ldots(4)$$

the summation now including every element of the circuit, and E, the e.m.f. of the generator, being defined as its *open-circuit voltage*. The introduction of this new defined magnitude is not really necessary, but it has proved to be useful in practice. We can establish that the magnitude so defined is additive in series connection and capable of either sign, so that the general expression for Kirchhoff's Second Law becomes

$$\Sigma IR = \Sigma E.$$

The unit of e.m.f. is, of course, the same as that of voltage, and the well known Weston standard cell can be regarded as a standard of either e.m.f. or voltage, though voltage is unquestionably the more fundamental property. Another point to notice is that we have extended the notion of resistance; for the property r is not measurable by the process by which the magnitude *resistance* is primarily

defined. Experimental justification for identifying the two properties can be obtained later by showing that correlation can be established between them by means of measurements with alternating current.

§ 10. ROUTINE METHODS

Having now established the basic laws of direct-current networks we can use them in order to establish all the bridge networks and potentiometer circuits and the great variety of operations by means of which measurements of current, resistance, voltage and e.m.f. are made in everyday practice. We can now use Ohm's Law for example for the purpose of measuring voltage in terms of resistance and current when that procedure proves to be more convenient than the one described. It should be noted here that when we say that two different processes serve to measure *the same magnitude*, we mean that experiment has shown that the two processes give the same result apart from a scale factor.

Routine methods differ very considerably from those we have described, mainly because the operations needed in order to establish a law are not necessarily those most convenient for subsequent applications of that law. When once our basis has been firmly established, our procedure is mainly dictated by consideration of economy of labour and the consistency of our measurements from day to day and year to year. The ease of construction of the required apparatus and its permanence become of dominating importance. In practice we find that a standard series of resistors is the most easily constructed, most permanent, and most complete as regards range and fineness of subdivision of all electrical standards. The general tendency is, therefore, to base most operations on measurements of resistance. Thus the operations we have described may appear at first sight as unfamiliar and not representative of actual practice. The practical application of Kirchhoff's First Law may take the form of a search for " leakage " or a " stray current " in order to explain some discrepancy in the observations. The experiments we have described are admittedly not frequently made deliberately with the idea of establishing a law. We think, however, that the experimentalist will recognize them as the logical equivalent of the mass of indirect experimental evidence for the truth of these laws which he has acquired in the course of his work.

We have now reached a stage at which it becomes necessary to establish relations between the basic electrical magnitudes and those of other branches of physics. Probably every experimentalist will be prepared to take for granted length, mass and time as three independently measurable magnitudes. We find, moreover, that over long periods of time our measurements of these three quantities show greater accuracy and consistency than those of any other physical quantities, and we may well therefore regard them as fundamental. We shall here also take for granted other non-electrical magnitudes that are usually held to be established in terms of length, mass and time by well known relations. The means by which these relations can be established is a separate subject.

§ 11. FARADAY'S LAWS OF ELECTROLYSIS

Among the simplest of the laws that we can now establish are Faraday's Law for the various voltameters and electrolytic cells. We find that if I denotes the

current in the cell, the mass, W, of any element liberated in time t can be represented by the relation

$$W = \beta . I . t. \qquad \qquad \ldots\ldots(5)$$

The constant of proportionality β is found to be characteristic of the element, and is therefore established as a derived magnitude, measurable for the various elements in terms of mass, current and time. We may fix the unit of β by making the scale-factor unity correspond with existing units of W, I and t, as when we measure β in " grams per ampère-second ". Alternatively we may fix the unit of current by assigning a definite value to β for silver (which is the element for which the most accurate experiments have been made), when the scale factor is unity and the units of W and t are those already chosen. The existing " International ampère " was defined in this way at a time when the silver volta-meter was considered·to be the best standard for current measurements, being easily available at all times and in most places, and giving very consistent results within a prescribed range of conditions.

§12. POWER DISSIPATION

When we study the filament lamp and the neon lamp of our original circuit we find no simple law relating either the luminous flux or the heat developed to the current or voltage alone, but calorimetric measurements show that heat is produced by the lamps at a rate proportional to the product of their current and terminal voltage. Further experiments with other circuit components show that this law is quite general, and that the proportionality constant is no more·than a scale factor. The units are therefore chosen so that we may express the relation in the simple form

$$P = VI. \qquad \qquad \ldots\ldots(6)$$

By the principle of the conservation of energy the dissipation of energy in a circuit must be accompanied by a uni-directional change in some part of the system, which thereby loses that energy. In our circuit the change is found to be chemical and to occur in the battery. We may say that the battery loses chemical energy equivalent to the thermal energy gained by the rest of the circuit, and that the difference between gain and loss of energy corresponds with the difference in the sign of V and, therefore, of VI for the two parts of the circuit.

§13. THE LAWS OF RESISTORS AND CONDUCTORS

Any object that may form part of an electric circuit, so that it may be characterized by a finite current, may be called a *conductor*, but we do not find it possible to determine a definite conductance and resistance for every conductor. Those for which definite values may be found obey Ohm's Law, and may be called *resistors*; but others, like the neon lamp, do not obey Ohm's Law even approximately: they are often called non-linear conductors, since they are characterized by a non-linear relation between current and voltage. It may be remarked that the law of power dissipation holds for all conductors, both linear and non-linear.

As soon as the measurement of resistance and conductance has been established, it becomes possible to establish laws relating these properties to temperature.

The experiments are complicated by the dissipation of power in the resistor, which restricts the range of current that can be employed in the detector circuit, but within limits they are practicable. When Ohm's Law is applied to any such resistor, due account must be taken of the variation of its resistance with temperature, and if this is done the law is found to hold good. On the other hand, the relation between voltage and current is non-linear, except for very small currents, on account of the variation of the temperature, and therefore of the resistance, with current. The filament lamp is therefore an example of a non-linear conductor or resistor which obeys Ohm's Law, while the neon lamp is a non-linear conductor which does not obey Ohm's Law. The most precise experiments show that all resistors are non-linear if we cover the widest possible range of current and voltage, and any measured value of resistance or conductance is therefore of significance only within a limited range of current and voltage.

We can also establish relations between the resistance and conductance of conductors of given material and their size and shape, but we must first notice that we have so far assumed that our circuits and, therefore, their components are " linear " in a sense quite different from that just mentioned. We have regarded the circuit as defined by a succession of points or terminals, such that voltmeters connected across these points, and ammeters connected between them, give definite indications, characteristic of the circuit or its elements. The circuit is therefore linear in the sense that it can be represented by a line having the same value of current at every point on it, and that the location of any voltmeter terminal on the circuit is as definite as that of a fixed point on this line. We immediately think of thin wires as providing the simplest method of achieving the conditions described, and although we have mentioned neon lamps and voltameters, which are certainly not thin wires, we must suppose that they have been joined by thin wires, and that each terminal is some definite point on such a wire. We think of the wire as providing a linear path for the current at the crucial points. When we come to deal with conductors of large cross-section, we think of current as entering terminal areas rather than terminal points, and we find that, by fixing one voltmeter terminal at some definite point on such a conductor and using the other as a probe, we can trace on the surface of a conductor lines of constant voltage called, for reasons to be given later, "equipotentials". Thus a terminal is now regarded as a device by means of which a voltmeter can be connected to an equipotential rather than a point. The experiments of high precision, by means of which the relations between the electrical properties of conductors and their geometrical properties have been established, have been almost entirely confined to straight conductors of constant cross-sectional area. Experiment shows that when these constitute parts of circuits, the equipotentials lie in parallel planes, and that the following law holds good :

$$\frac{I}{V} = \sigma \frac{A}{l}, \qquad \qquad \dots\dots(7)$$

where I denotes the current in the conductor, V the voltage corresponding with any two equipotentials separated by a distance l, A the cross-sectional area of the conductor, and σ is a constant for any one material, but varies with the

material. For completeness we should include a scale factor, but since I/V, by Ohm's Law, measures a conductance, the unit of which is fixed by other considerations, and the units of A and l are also fixed by other considerations, it is usual to make the scale factor unity for this particular set of units, thereby fixing the unit of the derived magnitude σ, conductivity.

By inverting the equation (7) we can show that the same experiments also establish a second derived magnitude, resistivity ρ, which becomes the reciprocal of σ, if the same convention is followed in fixing the unit. We have

$$-\frac{V}{I} = R = \rho\frac{l}{A}. \qquad \ldots\ldots(8)$$

It is obviously desirable that we should be able to deal with conductors of other shapes, and we therefore try to establish laws of greater generality. At this stage we find it necessary to introduce purely theoretical conceptions. We regard the conductor as the field of a vector $\mathscr{E}$ which is equal in magnitude but opposite in direction to the gradient of a scalar function v, the potential. It follows that the line integral of this vector between any two points is independent of the path followed, and simply equal to the difference of v for the two terminal points, and, moreover, for any closed path the line integral is zero. The function clearly has properties akin to those of voltage, and we can regard it as a generalization of voltage if it satisfies the condition that the difference of potential for any two points between which a voltage can be measured is equal to that voltage. As a generalization of current we introduce another vector $\mathscr{J}$, the "current density", which satisfies the two conditions: (a) that its flux across any terminal area for which a value of current can be measured is equal to that current, and (b) the "law of continuity",

$$\text{Div } \mathscr{J} = 0, \qquad \ldots\ldots(9)$$

which implies that the flux of $\mathscr{J}$ across any closed surface is always zero, and which we may regard as a generalization of the law that the current is the same along any linear circuit. We now write as a generalization of the law established for straight conductors of uniform cross-section,

$$\mathscr{J} = \sigma\frac{dv}{dn} = \frac{1}{\rho}\frac{dv}{dn}, \qquad \ldots\ldots(10)$$

where n is the normal to an equipotential surface, σ is the conductivity, and ρ the resistivity of the medium at the relevant point. This relation implies that the lines of flow of $\mathscr{J}$ are everywhere normal to the equipotential surfaces and therefore identical with the lines of flow of $\mathscr{E}$. These theoretical relations are postulates, for the truth of which we can never obtain direct experimental evidence; for $\mathscr{J}$ and v, and even $\mathscr{J}\,ds$, the current crossing a small area ds, are not measurable: it is always impossible to insert an ammeter into the conducting medium without altering the conditions appreciably. The postulated relations have, however, been found to lead to all the relations that have been established experimentally for conductors of various shapes. As we have seen already, equipotential lines can be traced experimentally on the surfaces of solid conductors, and in liquid conductors, equipotential surfaces having given voltages from one electrode may be traced out in a similar way. In electrolytes, moreover,

it can be shown that the thickness of deposited metal has some relation to current density. The postulated quantities are therefore not entirely without experimental significance, though the experiments in which they are significant are only of low accuracy. We shall later find applications for these conceptions in other branches of the subject.

§14. AMPÈRE'S LAW

We have seen that the basic laws in any branch of the subject are those which have been established by, or which govern the operation of, the instruments of highest precision available for work of that kind. It follows that the fundamental law of electrodynamics is the law of operation of the current balance or electrodynamometer, which is, broadly speaking, Ampère's Law, although we shall find it desirable to express it in terms different from those actually employed by Ampère.

Very simple experiments are sufficient to show that the mechanical force between two circuits is proportional to the product of the currents they carry; but the relation between the constant of proportionality and the geometrical properties of the circuits is so complicated that its discovery by Ampère was described by Clerk Maxwell as one of the greatest achievements in physical science. Ampère, by making inspired guesses based on the examination of circuits of simple form, found the form of the constant almost as soon as he discovered the proportionality. Subsequent work has served only to confirm the law with higher and higher precision, and to define the conditions in which it holds good. Ampère expressed his law in terms of the forces between circuit elements, and thereby introduced an ambiguity that has been the foundation of many paradoxes. But this ambiguity can be removed by translating his law into terms of the shapes and geometrical relation of complete circuits. It then becomes

$$F = kI_1I_2(-\partial N/\partial q), \qquad \ldots\ldots(11)$$

where F is the force tending to increase a coordinate q; I_1, I_2 are the currents in the two circuits; and N is defined as

$$N = \oint_1 \oint_2 \frac{dl_1 . dl_2 . \cos\theta}{r}, \qquad \ldots\ldots(11.1)$$

where dl_1, dl_2 are infinitesimal elements of the circuits 1, 2; r the distance, and θ the angle, between them. Thus N and $(-\partial N/\partial q)$ can be determined by geometrical measurements on the circuits.

Equation (11) is limited by its form to strictly linear circuits; but it can be extended in an obvious manner to any pair of circuits that can be regarded as each made up of linear filaments, rigidly connected, the proportion of the circuit current passing through each filament being known. But there is another limitation, usually expressed by the statement that the circuits must be composed of, and the space between them occupied by, a medium that is non-magnetic. We are not yet in a position to define " non-magnetic ", and when we are we shall find that none of the materials usually present in the experiments to which (11) is applied are, strictly speaking, non-magnetic. However, a sufficient condition

for (11) to be true can be stated definitely; it is that every medium in the neighbour-hood of the circuits should be either air or equivalent to air, equivalent meaning that, if any part of the medium that is not air is replaced by air, or *vice versa*, no appreciable change in the force between the circuits is produced. Equivalence thus depends on what changes are appreciable—that is to say, on the sensitivity of the measurement of force.

In this form and with these limitations the law (11) has been confirmed by many investigations with an accuracy that has steadily increased down the years. The most recent experiments at the National Physical Laboratory and the National Bureau of Standards agree and confirm the law to about 2 parts in 100,000.

In most treatises k is written μ, and treated as a derived magnitude whose value may vary. The basis of this practice is the belief that, if the air and its equivalents used in current balances were replaced by a medium not equivalent to air, and therefore magnetic, then, so long as the medium were uniform, (11) would still be true if k were allowed to take a different value. No experiments comparable in accuracy with those in air have been made in any other medium; and it is impracticable to use in such experiments the most important magnetic materials, which are solid and would therefore hamper greatly the measurement of force. Accordingly the direct evidence that k is a derived magnitude, and that it ought to be split into two factors $S\mu$, is of a quite different nature from that for (11), subject to the condition stated above. On the other hand, there are good theoretical grounds, which will be considered later, for thinking that k may be a derived magnitude. In these circumstances it is convenient to take advantage of the liberty noted in §2 and to write

$$F = S\mu I_1 I_2 (-\partial N/\partial q), \qquad \ldots \ldots (12)$$

leaving open for the time the question whether μ is a derived magnitude. It should be noted that (12) still requires that the medium about the circuits should be uniform. If it is not uniform in the sense that an interchange of the medium in one region with that in another might produce a change of the force, a law of the form (11) or (12) can give no account of the matter; for they contain no term that depends on the position of any material that does not form part of the circuits.

The law of the current balance has now been established with such accuracy that it can be used to define the unit of current. The law is (11) for the special case in which $I_1 = I_2 = I$, and we may write it

$$F = S\mu_0 I^2 (-\partial N/\partial q), \qquad \ldots \ldots (13)$$

the factor μ_0 serving to remind us that we are limited to an air-equivalent medium. The ampère has been defined by assigning to $S\mu_0$ in this law the value 10^{-2} when F is measured in dynes, or 10^{-7} when F is measured in m.k.s. units. The constant μ_0 in (13) is commonly termed the permeability of free space; it is perhaps worth noting that the relations implied by this term lie quite outside the basic facts that serve to determine the unit of current.

§15. MAGNETIC FIELD STRENGTH

Ampère, in his investigation of Oersted's well known discovery, also enunciated the law of the " moving magnet " indicator. Instruments of the type so named

have been found to be incapable of the same accuracy as the current balance, and the law has therefore never been established with high accuracy. It has, however, been found to be consistent with the available evidence, and we shall therefore accept it as a hypothesis. It can be written in the form *

$$T = mSI \oint \frac{dl \times r_1}{r^2} = SmIG, \qquad \qquad \ldots\ldots(14)$$

where T represents the torque on a small compass needle, about its axis of rotation, arising from the presence in its neighbourhood of a circuit carrying a current I; S is the scale factor; dl represents in length and direction any element of the circuit; r_1 is a unit vector in the direction from this element to the magnet; $\times$ represents a vector product; r is the distance between the element and the magnet; the integration is taken completely round the circuit; and m is a derived magnitude characteristic of the needle, its direction always perpendicular to the axis of rotation.

The form of the law requires that the dimensions of the magnet should be small compared with any r, in order that G, the integral, should be definite. A limitation is also imposed on the nature of the medium surrounding the magnet and circuit; a sufficient, but perhaps not a necessary, condition is that the medium should be air or its equivalent in the sense defined above.

The experimental evidence for (14) is scanty and usually indirect. Thus, though the law is implied in the well known experiments with sine- and tangent-galvanometers, those experiments do not enable m to be determined in terms of current, torque and the geometrical magnitude G; for the torque is measured in terms of that which the needle experiences in the absence of a current circuit, and it is assumed that this torque is also proportional to m. In particular, there is very little evidence that m is independent of I and G, as the law requires; indeed, there is indirect evidence that, for large values of I, m is not independent of I. However, such doubts are of little consequence, because m is not an important derived magnitude.

The real importance of (14) arises from its implication that two circuits for which IG is the same have the same effect on any compass needle. This law proves to be a special case of one more general; the torque on a compass needle is only one of several effects that can be observed in the neighbourhood of current circuits, e.g., the deflection of a cathode ray and the production of an e.m.f. in a rotating coil. All such effects (so long as the media are non-magnetic) are determined by IG in the sense that two circuits for which IG is the same produce the same effect, and that two circuits whose IGs are equal and opposite compensate each other and produce no resultant effect.

The experimental evidence for this statement is much stronger than that for (14). Some of it arises from an intimate mathematical connection between G and N (it will be stated explicitly in a later section), which was used at the National Bureau of Standards in some of the experiments on the current balance. Instead of determining N wholly by geometrical measurements, they determined

---

* This form is taken from G. P. Harnwell, *The Principles of Electricity and Magnetism* (McGraw-Hill, New York and London, 1938). This book seems to be representative of the working ideas of the experimentalist of today.

an important part of it as the ratio of the *G*s for the fixed and moving coils when these were placed suitably relative to each other; this they did by determining the ratio of the currents in the two coils at which the resultant effect on a compass needle vanished. The fact that they could thereby obtain consistent values of *N* for different pairs of coils, and that this method of determining *N* led to results agreeing with those of the National Physical Laboratory, who relied entirely on geometrical measurements, is very complete evidence for the significance attributed to *IG* as well as for (11).

Another part of the evidence for the significance of *IG* rests on the " Schuster magnetometer ", which is now recognized as the most accurate instrument for the measurement of the earth's magnetic field. (We are not yet in a position to define this quantity; all that we need to know about it here is that it is something that can be compensated by a current circuit.) This instrument consists essentially of a current circuit constructed so that the value of *G* for a small region fixed relative to it can be determined with high accuracy, some form of field detector being mounted within this region—in practice a small compass needle, a rotating coil, or a vibrating circuit mounted at the centre of a Helmholtz coil-system. The field within the small region is measured by observing what current must flow in the circuit in order that the detector shall indicate zero. Experiments with this instrument establish with high accuracy that the compensating current is inversely proportional to *G* at the relevant point.

These facts lead us to define a magnitude *H*, characteristic of points in the neighbourhood of a linear circuit, by

$$H = S.I.G. = S.I. \oint \frac{dl \times r_1}{r^2}. \qquad \ldots \ldots (15)$$

We insert the scale factor merely to allow for changes of unit. In accordance with general practice we shall call *H* the " magnetic field strength due to the current circuit ". Notice that it is not a magnitude established by experimental facts, but one laid down by definition; its importance depends on the fact that the quantity so defined proves to be significant in a wide range of experiments of the kind mentioned above. The definition can be extended to a non-linear circuit, so long as it can be regarded as a collection of linear circuits, by making *H* (which is a vector) mean the vector sum of components each associated with one of the linear circuits. In this extended meaning *H* retains its significance.

Purely mathematical reasoning is sufficient to show that the vector *H*, defined by (15), is the negative gradient of *Iω*, where *ω* is the solid angle subtended by the circuit at the point to which *H* refers; and *Iω* is therefore recognized as the magnetic scalar potential. It follows that the line integral of *H* around any closed path that does not link a circuit is zero, and that the line integral around a closed path that links *n* times in the same sense a circuit carrying a current *I* is given by

$$\oint H.dl = S.4\pi nI. \qquad \ldots \ldots (16)$$

We may remark that the two units of *H* that are in practical use arise naturally from equations (15) and (16). The oersted, the unit of the c.g.s.*μ* system,

was chosen so as to simplify (15) by making $S = 1$. The "ampère-turn per metre", the unit of the m.k.s. system, was chosen so as to simplify (16) by making $S.4\pi = 1$. There is no difference in principle between the two choices, and usage alone can show which is the more judicious.

Finally a word about magnetic media. If the medium surrounding the circuit is magnetically heterogeneous, (14) ceases to be true; circuits with the same IG are not in generally equivalent, so that H loses its significance if it is defined by (15), and there is no reason to adopt that definition. Whether in these circumstances there is any other quantity which plays in any sense the same part as H plays when the medium is non-magnetic is a question that must wait until a later section.

If the medium is uniform, but not equivalent to air, the scanty evidence available suggests that circuits with the same IG are still equivalent. Since (15) merely expresses this equivalence, there is no need to introduce a factor μ into it; whether it is desirable to do so, and what modification, if any, is required in (14), are again questions that must be postponed.

At this stage we may regard all the basic principles of circuits under direct current conditions as having been established. We have considered so far only currents, voltages and circuits that are invariable with time, and mainly materials that are non-magnetic. In subsequent parts we propose to deal with varying currents and voltages, which lead to the conceptions of inductance and capacitance, and so to the principles of magnetism, electrostatics and electromagnetic waves.

In concluding the first portion of our enquiry we would emphasize that we are not trying to establish a new system of electrical measurements, but merely to discover the logical basis of the system which is in fact universally practised. It follows that all the relations that we develop will be familiar ones, and the differences between our treatment and that of the standard treatises may appear trivial, especially in this first portion. The detail given is, however, necessary for logical completeness and to establish the method. The differences will become more significant in electrostatics and magnetism, and it is from these branches of the subject that we hope to remove many of the difficulties that frequently disturb the experimentalist. To the mathematician there are many possible ways of developing the relations of electromagnetism, and they are all exactly equivalent in logic and differ only in elegance. The experimentalist has less latitude; his practice is dictated by stubborn facts that defy representation in equations. The one course he is compelled by circumstances to follow is more important to him than the more elegant courses he might have followed in a better world. This is the excuse we offer to those who find our treatment clumsy.

LOW-ORDER MULTIPLE-BEAM
INTERFEROMETRY

By S. TOLANSKY,

University of Manchester

ABSTRACT. A discussion is given of the various factors affecting the intensity and sharpness of the multiple-beam Fizeau fringes and fringes of equal chromatic order used for the study of surface topography. The effects of absorption in the silver film, phase condition, linear displacement of the beams, finite size of source, departure from parallelism and source line width are considered. It is shown that Fabry-Perot fringes, multiple-beam Fizeau fringes and fringes of equal chromatic order have all the same fringe shape if certain conditions are fulfilled.

§1. INTRODUCTION

IN a group of earlier papers (Tolansky, 1944, 1945 a, b) it has been shown how multiple-beam interference methods can be applied to the study of surface topography. It is the purpose of this paper to discuss the factors affecting the definition of such fringes. The multiple-beam interference is produced by depositing highly reflecting thin silver films upon the two surfaces between which interference takes place. The fringe definition is affected mainly by the reflection coefficient, the geometrical conditions and the nature of the light. Each of these features will be considered in turn. The discussion is simplified by treating the case of interference in an air film ($\mu = 1$) and by first reviewing the case of interference between plane parallel silvered surfaces (i.e. Fabry-Perot interferometer).

§2. INTENSITY OF THE FRINGES

For multiple-beam interference between two parallel plane silvered surfaces, distant t in air, with reflecting coefficient R, the well-known Airy summation formula shows that the resulting fringes have an intensity distribution I given by

$$I = \frac{I_{max}}{1 + F \sin^2 \delta/2},$$

in which $F = 4R/(1-R)^2$ and δ, the phase difference between successive beams, equals $2\pi/\lambda . 2t \cos \phi$; the symbols here have the usually accepted meanings, and I_{max}, which is the intensity of the fringe maximum, equals $T^2/(1-R)^2$, where T is the fraction of light transmitted by each silver film. If there is no absorption, $I_{max} = 1$, but if a fraction A is absorbed by each film, then as $T + R + A = 1$, the value I'_{max}, involving absorption, is $I'_{max} = \left(\frac{1}{1 + A/T}\right)^2$, Airy's formula showing that it is desirable to have R as high as possible in order to produce the sharpest fringes. It is possible with silver to make R as high as 0·95 (in

the red end of the spectrum), but such a silver film has so high an absorption that it is useless for interferometry. Indeed, the *absorption* is the important factor which limits the film thickness that can be tolerated, and hence the reflectivity. Strong (1940) has given the values of T and R for evaporated silver films, up to $R=0\cdot89$, and the curves are such that it is permissible to extrapolate these slightly to include $R=0\cdot90$, this value being convenient for calculation. From these data A is simply obtained, and thus the quantity I'_{max} can be evaluated. The result of this simple calculation is shown in figure 1, which gives the intensity of the fringe maximum as a fraction of the incident intensity for different R-values.
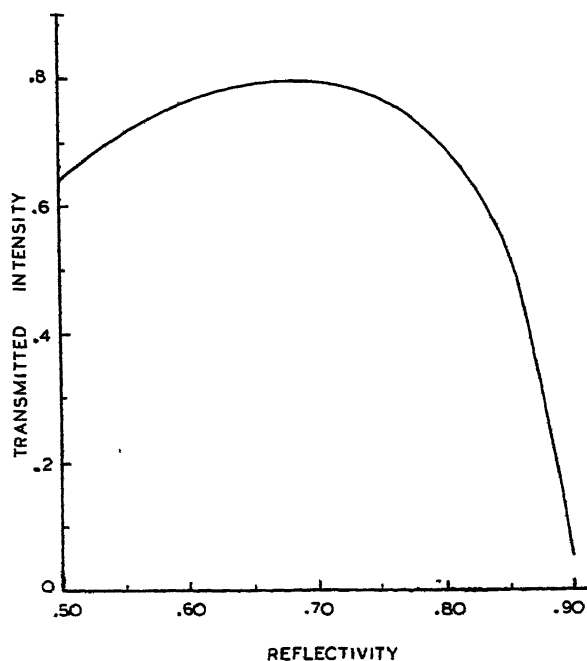
REFLECTIVITY

Figure 1.

This curve only obtains if particular attention is paid to purity conditions, for slight impurities seriously increase the value of A.

The fall in I'_{max} beyond a reflectivity of $0\cdot80$ is very rapid, and at $R=0\cdot90$ only some 5% of the incident light is transmitted. This curve shows why it is rarely possible to exceed $R=0\cdot87$ when using a Fabry-Perot interferometer for hyperfine structure studies. An increase of R from $0\cdot85$ to $0\cdot90$ improves the resolving power by 60%, but reduces the intensity by a factor of more than 10. Intensity considerations thus restrict the reflectivity that can be employed.

[*Note added in proof,* 2 *September* 1946: Measurements recently made in this laboratory enable figure 1 to be extended to $R=0\cdot94$, at which point I'_{max} is only $0\cdot75$% of the incident intensity.]

Fortunately in topographical studies very intense sources are available, e.g. high-pressure mercury arcs for Fizeau fringes and powerful carbon arcs for white-light fringes. This permits of the employment of much denser silverings

than are practical in hyperfine structure spectroscopy, and the loss of 95%, or even more, of the incident light can thus be tolerated, even if the fringes are observed by a microscope of × 100 magnification, which is at times the case. Reflectivities of 0·90 and more can thus be employed. With such a high reflectivity, some 70 to 80 multiple images of a lamp filament can be counted, although the tail images are weak. One can certainly assume that at least 60 beams will effectively combine in the multiple-beam interference.

From Airy's formula the half-width of the fringes (width at half intensity) as a fraction of an order is $2/\pi F^{\frac{1}{2}}$, which, for $R = 0.90$, equals 1/30 and for $R = 0.94$ is 1/50 of an order. Fringes have been obtained which approach the latter case.

§3. VARIETIES OF MULTIPLE-BEAM INTERFERENCE

It can be shown that under specified conditions the Airy formula holds fairly closely also for surfaces which are slightly inclined instead of parallel. By writing

$$I = \frac{I_{\text{max}}}{1 + F \sin^2\left(\dfrac{2\pi t \cos \phi}{\lambda}\right)}$$

it is clear that identical fringe shapes can arise from separate independent variations in either ϕ, t or λ providing conditions are such that all the effective beams add up and the phase differences are constant. It will be shown later that this condition can be approached, and assuming it to hold, it follows that four distinct types of multiple-beam fringes can be formed, *and in each case the fringe-intensity distribution is the same*, since a similar formula holds. The following table shows the fringe types possible.

Light	Constant quantity	Fringe type	Name	Equivalent filter
Monochromatic (λ constant)	t	Equal inclination	Fabry-Perot	Angular
	ϕ	Equal thickness	Fizeau	Linear
White (λ variable)	ϕ	Equal t/λ	Equal chromatic order	Wave-length
	t	Equal $\dfrac{t \cos \phi}{\lambda}$	White light Fabry-Perot	—

The last case (white-light Fabry-Perot fringes) has no application to the present problems and will be disregarded, since these fringes have already been discussed elsewhere (Tolansky, 1945 c).

The fourth column is a crude yet useful alternative manner of classifying the fringes. In effect a Fabry-Perot interferometer is an *angular* filter passing only that light incident at angles close to ϕ values for which the order of interference is integral. In a similar way, with multiple beams, the Fizeau fringes of a wedge are such that the wedge is in effect a *linear* filter (grating). Finally,

with fringes of equal chromatic order (Tolansky, 1945 d) every point on the interference surfaces acts as a *wave-length* filter (more accurately, as a frequency filter. This property has been made use of for the production of a new type of interferometric colour filter, details of which will be published later).

§4. THE PHASE CONDITION

In the Fabry-Perot case, and in some fringes of equal chromatic order (e.g. those from a double-silvered slip of mica), t is locally constant and the emerging beams are parallel, with constant phase difference. The Airy summation is therefore obtained simply by collecting the parallel beams with a lens, since the summed effect then takes place at the principal focus.

In multiple-beam Fizeau fringes (and many cases of fringes of equal chromatic order) conditions are very different, for the fringes are effectively localized at the interference film and the paths of succeeding beams are not strictly in phase. This feature has been completely overlooked in the literature despite its key importance in obtaining high definition. Thus Benoit, Fabry and Perot (1913) employed multiple-beam wedge Fizeau fringes in their classical evaluation of the length of the metre. Adam Hilger Ltd. have used such fringes for at least 20 years for testing high-grade optical surfaces, and recently Rasmussen (1945) has described further applications, yet neither in these nor in any other case has the undoubted fundamental importance of the phase condition been recognized.

Let AB, CD be the two plane surfaces, at angle θ, then with normal incidence the paths of the 1st, 2nd and 3rd beams which meet to interfere at the point X are as shown in figure 2. It is clear that successive beams do *not* have a constant phase difference. It can be shown from the geometry that the path difference between the 1st and nth beams is closely enough $2nt - \frac{4}{3}n^3\theta^2 t$. For strict Airy summation the path difference should be $2nt$; thus the phases become progressively out of step and destroy the condition for summation unless the quantity $\frac{4}{3}n^3\theta^2 t$ is made small compared with λ.

Let it be assumed that the nth beam is a half-wave out of step and thus completely opposes the Airy summation. Taking this as a limit, then

$$\frac{4}{3}n^3\theta^2 t = \lambda/2.$$

If x is the number of fringes per cm. on the wedge surface,

$$\theta = \lambda x/2,$$

giving

$$t = \frac{3}{2n^3\lambda x^2}.$$

Suppose, for $\lambda = 5 \cdot 5 \times 10^{-5}$ cm., that n is taken to be 60, then

$$t = \frac{1}{7 \cdot 92 x^2}.$$

In some instances, fringes may be 1 cm. apart ($x = 1$), but more frequently they are 1 mm. apart ($x = 10$), or even 0·1 mm. apart ($x = 100$), the latter being the case when low-power microscopy is employed. Corresponding critical values for t are

Number of fringes per centimetre	1	10	100
Critical t (mm.)	1·26	·012	10^{-4}
d_{60} (mm.)	0·25	·025	·0025

The calculation gives only the approximate order of t for the case of $x = 100$, since t is then less than a light wave and the approximations used are invalid.

An important feature has emerged; for the Airy summation to be effective the values of t should be less than the values above; in other words *the distance between the surfaces must be as small as possible indeed—usually of the order of at most a few light waves.* This is the critical fact overlooked by others, for other important features depend on it also, as shown below.

§5. LINEAR DISPLACEMENT OF BEAMS

The phase retardation just calculated for the successive beams is dependent upon the linear displacement of the beams along the wedge surface. Thus it is advisable to use normal incidence since this displacement increases with increasing incidence.

In the case of the study of a complex topography it is clearly essential to view interference from beams which have scanned as small an area as possible. Figure 2
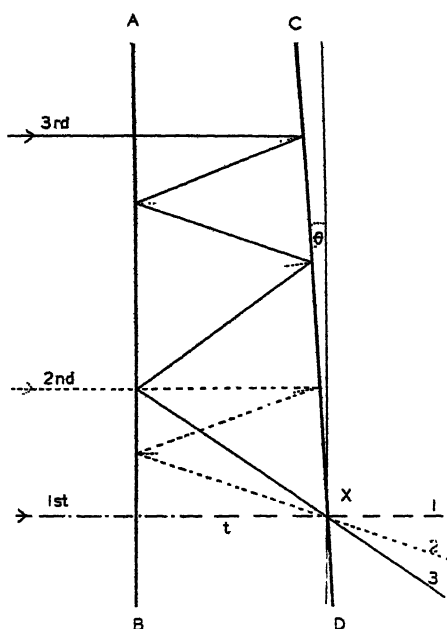
Figure 2.

shows that the higher-order beams come down from points successively further and further away from the first beam, and to a first approximation the linear separation on the surface between the 1st and nth beams is

$$d_n = 2n^2 t\theta = n^2 \lambda x t,$$

and substituting for t for the 60th beam (above) gives

$$d_{60} = \frac{1}{4x} \text{ mm.}$$

The values, for $x = 1, 10, 100$, are included in the previous table, third row.

Again on this ground, therefore, t must be as small as possible. If t is of the order of $1/1000$ mm., which is desirable, then d_{60} is of the same order, so that effectively for any feature resolvable by a low-power microscope all the beams scan within the limit of resolution.

§6. ERRORS IN COLLIMATION

Fabry (1922) first drew attention to the broadening effects in multiple-beam Fizeau fringes consequent on lack of parallelism in the incident beam arising from the finite dimensions of the "point" source. The following shows that with the values of t imposed of necessity by the phase condition, collimation defects are of no consequence.

The change in order dn produced by an angular deviation ϕ from the normal is

$$dn = \frac{2t}{\lambda} \sin^2 \frac{\phi}{2},$$

and as only small angles are under consideration

$$\phi = \sqrt{\left(\frac{\lambda \, dn}{t}\right)}.$$

If one is prepared to tolerate a change dn no more than $1/5$ the half-width, then, for $R = 0.90$,

$$dn = \frac{1}{150},$$

which gives the angle ϕ in degrees (for $\lambda = 5.5 \times 10^{-5}$ cm.) as

$$\phi^\circ = \frac{3.4 \times 10^{-2}}{t^{\frac{1}{2}}}.$$

Approximate values of ϕ for successive values of t are given below, and if a 10-cm. focal length lens is used, the corresponding diameters permitted for the source, D (in mm.), are shown in the third row.

t (mm.)	1	0·1	0·01	0·001
ϕ°	1/10	1/3	1	3
D (mm.)	0·2	0·6	2	6

Since $t > 0.01$ mm. is excluded, a source diameter of at least 2 mm. can be tolerated, often three times this. Clearly precise collimation is only critically important for the larger t values.

§7. HIGH-DISPERSION FIZEAU FRINGES

In a recent communication (Tolansky and Wilcock, 1946) a technique of "crossed" fringes has been described in which use is made of high-dispersion Fizeau fringes. In this technique θ is made as small as possible, only the residual values due to the surface configuration being permitted. With such an arrangement the fringes are very broad and interpretation is normally difficult, but this is made simple by the "crossing" technique. The high-dispersion fringes

then become a powerful method for the detection of small changes in thickness, of molecular dimensions, in accordance with the following: as θ diminishes, the Fizeau fringes broaden, and, with parallel or nearly parallel plates, this leads to a uniform tint of intensity determined by the value of t.   When t is an integral number of half-waves the tint is of maximum intensity and half-way between is a minimum, the value of the latter being only about $0\cdot25\%$ that of the maximum (for $R=0\cdot90$) and thus, in effect, zero.

The fringe shape is such that the maximum tint-change sensitivity to a change in thickness occurs if a value of t is adopted such that approximately half the peak intensity is within the field of view.   From the Airy formula we have the following: if δ is the change in phase from the n integral (maximum) position at which half-intensity occurs, then

$$\delta = \frac{2}{F^{\frac{1}{2}}}.$$

Supposing one can measure a change of 10% in tint intensity (a microphotometer could detect 5%), this corresponds to a phase δ' at which

$$I = 0\cdot45\,I_{max}\ (\text{or}\ 0\cdot55\,I_{max}).$$

Substitution gives

$$\delta' = \frac{2\sqrt{11}}{3F^{\frac{1}{2}}}.$$

From this the fraction of an order detectable δN is

$$\delta N = \frac{\delta'-\delta}{2\pi} \simeq \frac{1}{30F^{\frac{1}{2}}},$$

giving

$$\delta N \simeq \frac{1}{570},\ \text{for}\ R=0\cdot90.$$

This corresponds to only 5 A.

Thus by selecting the correct value of t, a change of ±5 A. in the topographical height can be measured.   Such a precision and sensitivity have not yet been attained by any of the other multiple-beam methods.   The method is, however, severely restricted in use and requires a microphotometer for general numerical evaluation.

§8. EFFECT OF SOURCE LINE-WIDTH

Since at normal incidence $n\lambda=2t$, the wave-length range between orders is $\Delta\lambda$, given by

$$\Delta\lambda = \frac{\lambda^2}{2t} \simeq \frac{0\cdot15}{t}\text{A. for}\ \lambda=5\cdot5\times10^{-5}\ \text{cm.}$$

The wave-length range for different t values is then as given below in the second row, in A.

t (mm.)	1	0·1	0·01	0·001
$\Delta\lambda$ (A.)	1·5	15	150	1500
$\delta\lambda$ (A.)	0·01	0·1	1	10
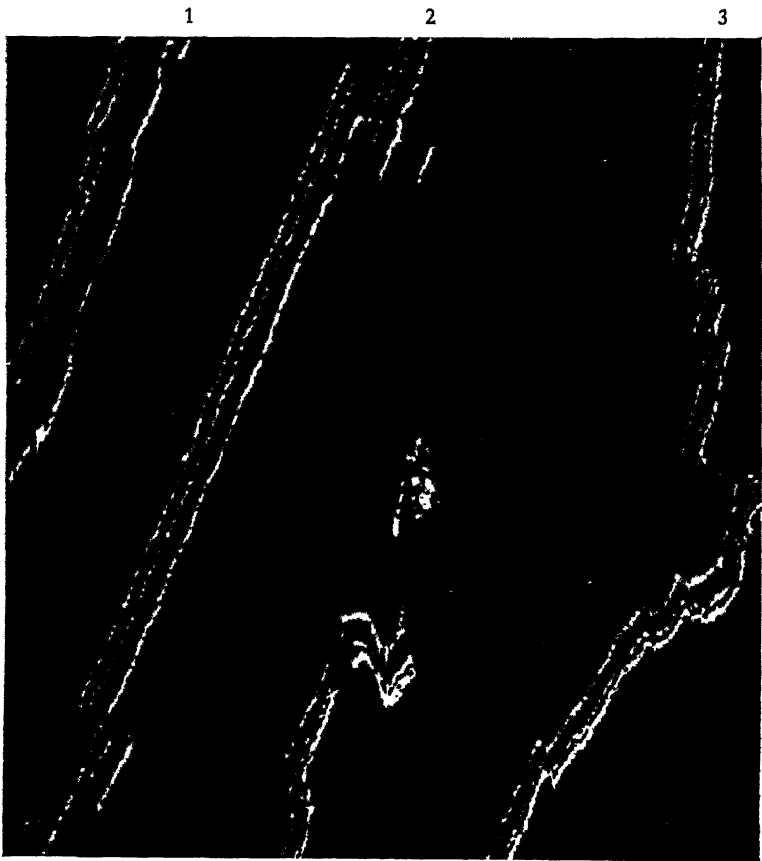
Figure 3. Fizeau fringes (mercury source) of a diamond face, area 2 sq. mm.

5760 A.

6230 A.

Figure 4. Fringes of equal chromatic order with calcite.

As in the treatment of the errors in collimation, it is assumed that 1/150 change in order is permissible. The line-width $\delta\lambda$ which can thus be tolerated is that shown in the third row. Since previous considerations limit t to be less than 0·01 mm., and preferably of the order 0·001 mm., a source line-width of 1→10 A. can be employed as compared with less than 0·01 A. needed for hyperfine structure. This fact is of the utmost importance, for this width tolerance permits the employment of an intense high-pressure mercury arc (usually, however somewhat under-run) with consequent great brilliance. The brilliance enables higher reflecting coefficients to be adopted in accordance with the intensity considerations of figure 1. Ultimately, then, much narrower fringes are obtained.

To sum up, by making t of the order of 0·001 mm., definition is greatly improved because

(1) a higher reflecting coefficient can be used;

(2) the phase condition is satisfied;

(3) linear displacement of multiple beams is reduced;

(4) errors in collimation are negligible;

(5) source line-width can be easily rendered innocuous.

As an example of the results of applying the principles discussed here, figure 3 in the accompanying plate shows the multiple-beam Fizeau fringes of 2 square mm. of a natural octahedron diamond face. 1, 2, 3 are successive orders of the green mercury line, the other fringes being the yellow doublet. The extreme sharpness is striking. The print is taken from a soft photographic plate.

§9. FRINGES OF EQUAL CHROMATIC ORDER

The former considerations as to the dependence of intensity on absorption, phase condition, linear displacement of beams and errors in collimation also apply to the white-light fringes of equal chromatic order given by a wedge. On these grounds, therefore, t should also be as small as possible for this type of fringe. Cases do arise, however, in which there is no violation of phase condition nor any appreciable linear displacement effect, e.g. doubly silvered mica film, in which locally the sides are frequently parallel, so that the conditions for the Airy summation are realized. Since the collimation defect is relatively trivial, it follows that as θ can always be reduced to a minimum by making the surfaces as near parallel as possible, then on these grounds alone fringes of equal chromatic order should be, and indeed are, somewhat sharper than the corresponding multiple-beam Fizeau fringes. In particular is this realized when higher magnifications are involved, for as already shown, in Fizeau fringes the phase condition is determined by θ^2, i.e. x^2, where x is the number of fringes per cm. This may be as much as 100, so as to give a reasonable fringe group in the field of view of the microscope. In fringes of equal chromatic order, the number and spatial distribution of fringes seen is *quite independent of the magnification used in projecting the image on to the slit;* thus θ can be made to approach zero, even for the highest of magnifications. This is an important advantage.

With a complex surface topography, there remains always a residual θ value due to the surface shape, hence the advisability of keeping t reasonably low.

However, the influence of t on fringe dispersion is much more important. Since the wave-number separation $\Delta\nu$ between fringes is $\Delta\nu = 1/2t$, and as a $\Delta\nu$ value of $1000 \rightarrow 5000$ cm.$^{-1}$ is desirable to show fine structure details with a typical Hilger medium quartz instrument, then t should be of the order $0\cdot005 \rightarrow 0\cdot001$ cm. Thus one arrives at a value identical with that needed to satisfy all the other conditions. Hence, again as with Fizeau fringes, t should be at most a few light-waves.

Since white light is used, line-width considerations are not involved. However, of much importance is the fact that a very intense white-light source is available in the high current-density carbon arc. This is so bright that silverings can be used which are definitely too thick even for the high-pressure mercury lamp used with Fizeau fringes. Thus again on this count fringes of equal chromatic order can be obtained sharper than Fizeau fringes. A striking example of the definition obtainable by attention to the details discussed is shown in figure 4 of the plate. These are fringes of equal chromatic order (positive) for a calcite crystal matched against a specially prepared flat. The wave-lengths are indicated in ångströms. The value of t is $0\cdot0015$ mm. The silvering is very thick, R exceeding $0\cdot90$ and probably approaching $0\cdot94$ judging from the fringe-width. Exposure was a matter of some 15 minutes despite the thickness of the silver. The fringes have a striking narrowness. In the picture can be seen a small cleaved-out strip only 60 A. deep, yet clearly many times the amount that could still be resolved (the lattice spacing in calcite is 6 A.: see Tolansky and Khamsavi (1946)).

REFERENCES

BENOIT, FABRY and PEROT, 1913. *Bur. Intern.*, Vol. II.
FABRY, 1922. *Revue d'Optique*, 1, 445.
RASMUSSEN, 1945. *Kgl. Danske Vidensk. Sel. Mat. Fys.*, 23, 4.
STRONG, 1940. *Modern Physical Laboratory Practice* (Blackie).
TOLANSKY, 1944 *Phil. Mag.*, 35, 120.
TOLANSKY, 1945 a. *Proc. Roy. Soc.*, A, 184, 41.
TOLANSKY, 1945 b. *Proc. Roy. Soc.*, A, 184, 51.
TOLANSKY, 1945 c. *Phil. Mag.*, 36, 236.
TOLANSKY, 1945 d. *Phil. Mag.*, 36, 225.
TOLANSKY and KHAMSAVI, 1946. *Nature, Lond.*, 157, 661.
TOLANSKY and WILCOCK, 1946. *Nature, Lond.*, 157, 583.

A TRANSFORMATION OF KNOWN ASTIGMATISM FORMULAE

By H. H. HOPKINS,

W. Watson and Sons, Barnet, Herts

ABSTRACT. The use of effective " centres " and " radii " of curvature permits the well-known formulae relating the tangential and sagittal centres of curvature of a wave-front element before and after refraction to be put into a form which yields formulae and computing schemes identical with those customarily used for computing the paraxial focus, magnification, and equivalent focal length. Simple expressions follow for the *T*- and *S*-magnifications and oblique equivalent focal lengths.

After being in use for a few months, the computing schemes have now been adopted as standard calculations.

IN figure 1, ACE'E is the optical axis of a lens system one of whose refracting surfaces, AP, is shown. Its pole is at A, its centre at C, and the radius AC=r. PE is a ray (continued) which is incident on the refracting surface at P, making an angle I with CP the normal at P. (For convenience the refracting surface is assumed to be spherical. The treatment can nevertheless be formulated for any surface of revolution.) The principal centres of curvature of an element
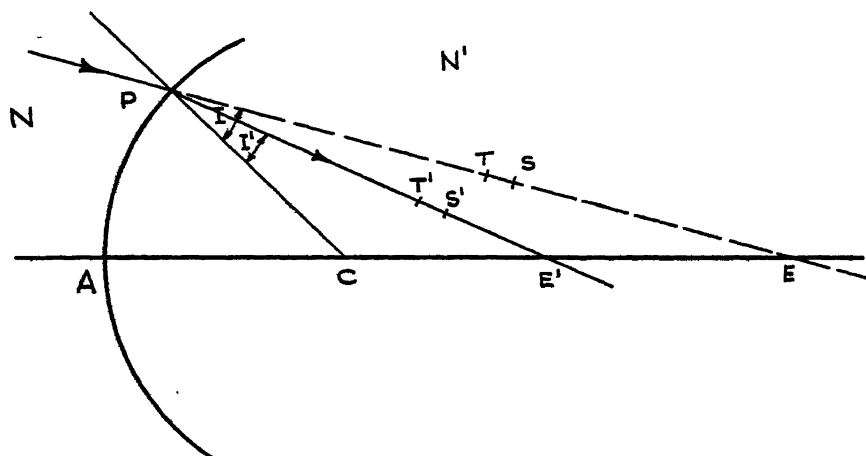
Figure 1.

of wave-front proceeding along this incident ray necessarily lie on the ray itself, since this is normal to the element of wave-front. The principal curvatures of the element of wave-front, assuming a ray in the meridian (tangential) plane, are known to be in the tangential and sagittal sections. The tangential section is the plane of the diagram in figure 1; and the sagittal section is a plane containing the ray and perpendicular to the plane of the diagram. Let T, S be the principal centres of curvature of the incident wave-front element; and let T', S' (lying

on PE′, the refracted ray) be these centres for the refracted wave-front element which proceeds along PE′. This ray makes an angle I' with CP. Writing PT $=t$, PT′ $=t'$, and PS $=s$, PS′ $=s'$, the well-known formulae connecting these lengths are

$$\frac{N'\cos^2 I'}{t'} - \frac{N\cos^2 I}{t} = \frac{N'\cos I' - N\cos I}{r}, \qquad \ldots\ldots(1)$$

$$\frac{N'}{s'} - \frac{N}{s} = \frac{N'\cos I' - N\cos I}{r}. \qquad \ldots\ldots(2)$$

Each of these formulae is analogous to the paraxial formula

$$\frac{N'}{l'} - \frac{N}{l} = \frac{N'-N}{r} \qquad \ldots\ldots(3)$$

relating the centres of curvatures before and after refraction of an element of wave-front which proceeds along the optical axis of the system. Equation (3) can be looked upon as the degenerate form, common to both (1) and (2), when $I = I' = 0$.

In lens computations, both for the purposes of calculating magnifications, equivalent focal lengths, and also for applying the sine condition and Herschel's condition, equation (3) is modified by the introduction of a paraxial incidence height at each surface y. (See Conrady (1929), whose notation is adopted except where stated to be otherwise.) The resulting formulae also afford useful check calculations.

In figure 2 is shown the diagram relating to an incident paraxial ray. The notation differs from Conrady's only in that $\overset{\wedge}{\text{PCA}} = \alpha$. The computing scheme may be summarized:

$$\frac{l_k - r_k}{r_k} \cdot u_k = i_k,$$

$$u_k + i_k = \alpha_k,$$

$$\ldots\ldots\ldots$$

$$\left.\begin{array}{l} y_k = r_k \cdot \alpha_k, \\ y_k = y_{k-1} - d_{k-1} \cdot u_k, \end{array}\right\} \begin{array}{l} \text{Check} \\ \text{values of } y_k \end{array}$$

$$\ldots\ldots\ldots$$

$$i_k' = i_k \frac{N_k}{N_k'},$$

$$u_k' = \alpha_k - i_k',$$

$$\left.\begin{array}{l} l_k' = \dfrac{i_k'}{u_k'} \cdot r_k + r_k, \\ l_k' = y_k / u_k', \end{array}\right\} \begin{array}{l} \text{Check} \\ \text{values of } l_k' \end{array}$$

$$\ldots\ldots\ldots\ldots$$

$$l_{k+1}' = l_k' - d_k,$$

$$\ldots\ldots\ldots$$

$$\left.\frac{(l-r)\cdot u}{(l'-r)\cdot u'} = \frac{N'}{N}.\right\} \begin{array}{l} \text{Additional} \\ \text{check} \end{array}$$

The scheme requires but simple, and well-known, modifications to meet the special cases $l = \infty$, $r = \infty$. The values of u_1, u_k' are used in finding the image size after the kth refraction, and in computing the coma as given by the sine-condition. They are also required in the computation of the change of spherical aberration consequent upon a change of object distance (see Hopkins, 1946).

Similar considerations can be applied to imagery along any ray, generally in practice a principal ray, which can then be looked upon as analogous to the axis. In what follows, it is shown that by introducing " effective " values of r, N, N', and a " paraxial " incidence height, there can be obtained formulae and computing schemes, of which that above is the degenerate case $I = I' = 0$. In each of the

:alculations (for the T- and S-centres of curvature) values of u_1, u_k' are given which lead to " oblique " magnifications M_T, M_S and to " oblique " equivalent 'ocal lengths F_T, F_S. The paraxial magnification and equivalent focal length igain appear as the degenerate cases $I = I' = 0$.

In figure 3 is shown an incident ray PT together with a neighbouring ray
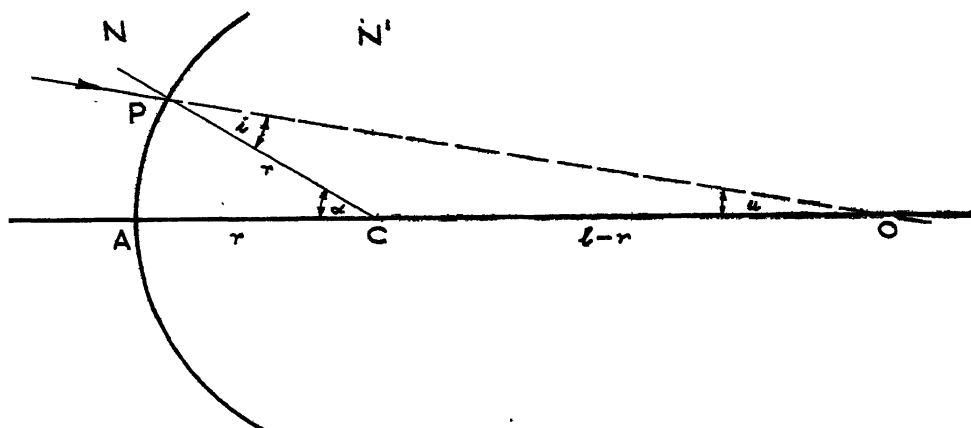
Figure 2.

$P_T T$ (in the tangential plane) which intersect at T, the tangential centre of curvature of the wave-front element proceeding along PT. $P_T T$ will be referred to as a T-ray and the point T as the T-focus; these are analogous to a paraxial ray and he paraxial focus. Draw a perpendicular from P_T on PT of length y; and draw
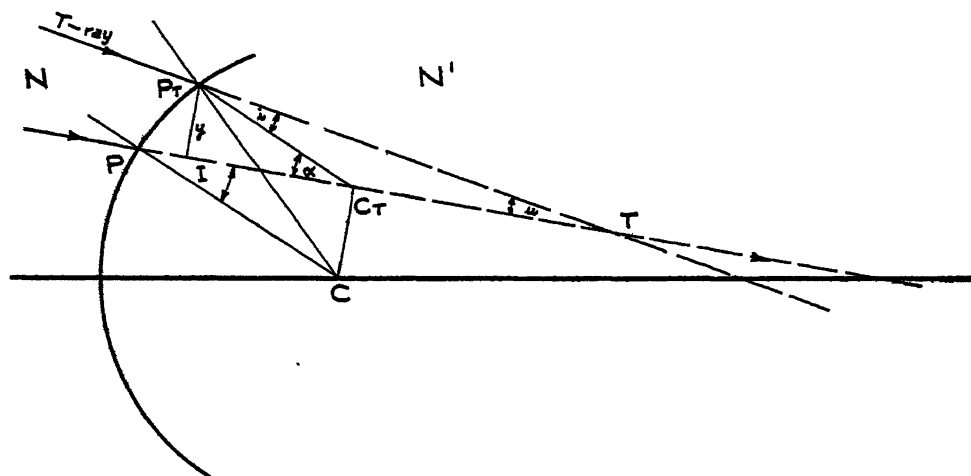
Figure 3.

CC_T perpendicular to PT to meet this ray in C_T. Then C_T will be looked upon s the effective " centre of curvature " of the refracting surface before refraction. The effective " radius of curvature " is $PC_T = r \cos I$. After refraction, C_T', ying on the refracted ray, and $PC_T' = r \cos I'$ are the effective " centre " and ' radius " of the surface. Join now $P_T C_T$: then the triangle $P_T C_T T$ corresponds

with the triangle PCO of figure 2 and a corresponding notation has been introduced. It will be seen that

$$u = \frac{y}{t}, \qquad \alpha = \frac{y}{r \cos I}, \qquad \alpha = u + i;$$

$$u' = \frac{y'}{t'} \qquad \alpha' = \frac{y'}{r \cos I'}, \qquad \alpha' = u' + i'.$$

Further, PP_T being small,

$$PP_T = \frac{y}{\cos I} = \frac{y'}{\cos I'},$$

so that

$$\alpha = \alpha', \quad \text{or} \quad u + i = \alpha = u' + i'.$$

Considering the triangle $P_T C_T T$,

$$i = \frac{t - r \cos I}{r \cos I} \cdot u$$

and, from the triangle $P_T C_T' T'$,

$$i' = \frac{t' - r \cos I'}{r \cos I'} \cdot u'.$$

It will be shown by (1) that the angles i, i' obey a paraxial " law of refraction " providing $N \cos I$, $N' \cos I'$ be written in place of N, N'.

Rearranging the terms in (1) gives, after multiplication of the left- and right-hand sides by $\dfrac{y}{\cos I}$, $\dfrac{y'}{\cos I'}$ respectively, gives

$$N' \cos I' \left\{ \frac{t' - r \cos I'}{r \cos I'} \right\} \cdot u' = N \cos I \left\{ \frac{t - r \cos I}{r \cos I} \right\} \cdot u$$

or

$$N' \cos I' \cdot i' = N \cos I \cdot i,$$

i.e., a " paraxial " law of refraction using " indices " $N' \cos I'$ and $N \cos I$.

It is required now to provide an expression for y_k in terms of y'_{k-1}. This is

$$y_k = y'_{k-1} - D_{k-1} \cdot u_k,$$

where D_{k-1} is the length along the principal ray between the $(k-1)$th and kth surfaces, and is analogous to d_{k-1} in the paraxial equations.

In figure 4, $P_s' S$ is a ray in the sagittal section close to the ray PS. S is the sagittal focus and $PS = s$. A similar notation with dashes applies to the conditions after refraction. PP_s is a line perpendicular to PS and is the " paraxial " incidence height y. Clearly, in this case, $y = y'$. From C draw CC_s perpendicular to PC and meeting PS in C_s. C_s is the effective "centre", and $PC_s = r/\cos I$ the effective " radius ", of the refracting surface. Introducing the notation used above,

$$u = \frac{y}{s}, \qquad \alpha = \frac{y}{r/\cos I}, \qquad \alpha = u + i;$$

$$u' = \frac{y}{s'}, \qquad \alpha' = \frac{y}{r/\cos I'}, \qquad \alpha' = u' + i''.$$

Considering the triangle $P_s C_s S$,

$$i = \frac{s - r/\cos I}{r/\cos I} \cdot u$$

and, for the triangle $P_s'C_s'S'$,

$$i'' = \frac{s' - r/\cos I'}{r/\cos I'} \cdot u'.$$

It will be shown by (2) that i, i' obey a paraxial law of refraction, in which N, N'' are the appropriate indices.

Rearranging the terms in (2) and multiplying by y gives

$$N' \cdot \left\{ \frac{s' - r/\cos I'}{r/\cos I'} \right\} \cdot u' = N \cdot \left\{ \frac{s - r/\cos I}{r/\cos I'} \right\} \cdot u$$

or

$$N' \cdot i' = N \cdot i,$$

i.e. a " paraxial " law using indices N, N'. As before,

$$y_k = y_{k-1} - D_{k-1} \cdot u_k.$$
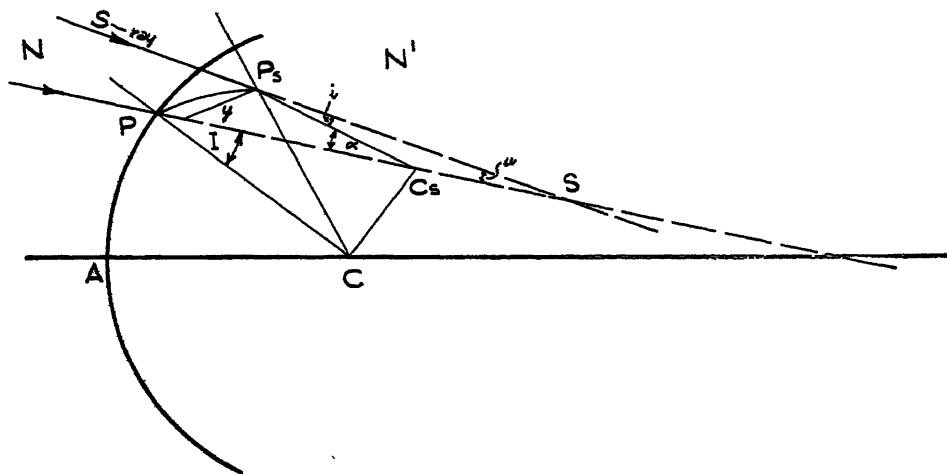
Figure 4.

Thus the computing schemes for the T- and S-calculations are:

$$\frac{t_k - r_k \cos I_k}{r_k \cos I_k} \cdot u_k = i_k,$$

$$u_k + i_k = \alpha_k,$$

$$\cdots \cdots \cdots$$

$$\left. \begin{array}{l} y_k = r_k \cos I_k \cdot \alpha_k, \\ y_k = y_{k-1} - D_{k-1} \cdot u_k, \end{array} \right\} \text{Check}$$

$$\cdots \cdots \cdots$$

$$i_k' = i_k \cdot \frac{N_k \cos I_k}{N_k' \cos I_k'},$$

$$u_k' = \alpha_k - i_k',$$

$$\cdots \cdots \cdots$$

$$\frac{s_k - r_k/\cos I_k}{r_k/\cos I_k} \cdot u_k = i_k,$$

$$u_k + i_k = \alpha_k,$$

$$\cdots \cdots \cdots$$

$$\left. \begin{array}{l} y_k = r_k/\cos I_k \cdot \alpha_k, \\ y_k = y_{k-1} - D_{k-1} \cdot u_k, \end{array} \right\} \text{Check}$$

$$\cdots \cdots \cdots$$

$$\alpha_k' = \alpha_k \cdot \frac{\cos I_k'}{\cos I_k},$$

$$i_k' = i_k \cdot \frac{N_k}{N_k'},$$

$$u_k' = \alpha_k' - i_k',$$

$$\cdots \cdots \cdots$$

$$t_k' = \frac{i_k'}{u_k'} . r_k \cos I_k' + r_k \cos I'_k, \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} \text{Check} \qquad s_k' = \frac{i_k'}{u_k'} . r_k/\cos I_k' + r_k/\cos I_k', \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} \text{Check}$$

$$y_k' = y_k . \frac{\cos I_k'}{\cos I_k},$$

$$t_k' = y_k'/u_k', \qquad\qquad\qquad\qquad s_k' = y_k/u_k',$$

. . . .

$$t_{k+1} = t_k' - D_k, \qquad\qquad\qquad s_{k+1} = s_k' - D_k,$$

·. . . . .

Additional checks :—

$$\frac{t - r \cos I}{t' - r \cos I'} = \frac{N'}{N} \qquad\qquad \frac{s - r/\cos I}{s' - r/\cos I'} = \frac{N' \cos I'}{N \cos I}.$$

The computing operations are identical with those of a paraxial calculation except for the calculation of y' from y, and α' from α. The special cases $t = \infty$, $s = \infty$, $r = \infty$ require simple modifications, again analogous to the paraxial formulae in these cases.

The transverse and longitudinal magnifications associated with imagery by T-rays are simply

$$(M)_T = \frac{N_1(u_1)_T}{N_k'(u_k')_T}, \qquad\qquad (M_{long})_T = \frac{N_1(u_1)^2_T}{N_k'(u_k')^2_T}.$$

Those associated with imagery by S-rays are

$$(M)_S = \frac{N_1(u_1)_S}{N_k'(u_k')_S}, \qquad\qquad (M_{long})_S = \frac{N_1(u_1)^2_S}{N_k'(u_k')^2_S}.$$

The tangential and sagittal equivalent focal lengths are given (when t_1, $s_1 = \infty$) by

$$F_T = \frac{y_1}{(u_k')_T}, \qquad\qquad\qquad F_S = \frac{y_1}{(u_k')_S}.$$

In these magnifications and equivalent focal length formulae the object and image elements are assumed perpendicular to the incident and emergent (principal) rays.

The forms of sine condition and Herschel's condition applicable to oblique imagery can be readily formulated by a method previously given for axial imagery (see Hopkins, 1946 a, b).

REFERENCES

CONRADY, A. E., 1929.   *Applied Optics and Optical Design* (London : Oxford University Press), Chapter I.
HOPKINS, H. H., 1946 a.   *Proc. Phys. Soc.*, 58, 92.
HOPKINS, H. H., 1946 b,   *Proc. Phys. Soc.*, 58, 100.

THE MECHANICAL PROPERTIES OF METALLIC SOLID SOLUTIONS

By F. R. N. NABARRO,

Royal Society Warren Research Fellow

ABSTRACT. The theoretical relation between the lattice strains produced by precipitation in a metal and the corresponding increase in hardness is extended to the case of lattice strains in metallic solid solutions. The elastic limit of a single crystal of a solid solution is calculated on the assumption that the crystal will slip when the applied external stress is equal to the mean value of the internal stress. This mean is taken over the length of a dislocation, assumed to be 1000 atoms. The estimate agrees in order of magnitude with the experimental observations. Similar considerations are applied to the hardness of polycrystalline solid solutions. The theory is extended to cases in which the increase of hardness produced by alloying is not large in comparison with the hardness of the pure solvent.

§ 1. INTRODUCTION

IT has long been known (Rosenhain, 1921) that the hardness of a metallic solid solution can be correlated with the distortion of the lattice caused by the presence of foreign atoms. A theory of precipitation hardening based on the concept of dislocations has already been given (Mott and Nabarro, 1940), and the aim of the present paper is to discuss the modifications in this theory which become necessary if it is applied to solid solutions in which the nuclei of strain are single atoms rather than aggregates of many atoms. These modifications are of two kinds. Firstly, it is no longer justifiable to estimate the strain in the matrix from purely elastic considerations of the atomic volumes of solvent and solute, and account must be taken of the binding forces between atoms of the solvent and of the solute. Secondly, even in solutions as dilute as 1%, the internal strains caused by the dissolved atoms change their directions over distances of a few lattice spacings, and it is to be expected that the opposed stresses acting on closely adjoining parts of a dislocation will largely cancel one another and so reduce the applied stress necessary to move the dislocation. It is apparently this effect which accounts for the great increase in mechanical strength produced by precipitation hardening.

§ 2. THE HARDNESS OF AN ALLOY CONTAINING A PRECIPITATED PHASE

According to the theory advanced by Mott and Nabarro, the elastic limit S of an alloy hardened by the precipitation of a second phase in roughly spherical particles should be given in order of magnitude by

$$S = E\epsilon f, \qquad \qquad \ldots\ldots(1)$$

where E is Young's modulus for the matrix (which to the order of accuracy required may be taken as the modulus for the pure solvent), f is the concentration

of the precipitated phase, and ϵ measures the misfit of the precipitate in the lattice of the matrix when the stresses in the matrix and in the precipitate are in equilibrium.

If this formula is applied to solid solutions, S and E retain their meanings unchanged, and f is the atomic concentration of the solute. The effective value of ϵ may be obtained directly from the difference of the measured x-ray lattice parameters of the solid solution and of the solvent. If the lattice parameter of the solvent is a and that of the solution is $a + \Delta a$, the fractional change in atomic volume due to the presence of the solute is $(1 + \Delta a/a)^3 - 1$. This fractional change is also given by $(1 + \epsilon f)^3 - 1$, and the relation between ϵ and Δa is thus

$$\epsilon f = \Delta a/a. \qquad \qquad \ldots\ldots(2)$$

§3. INFLUENCE OF THE SMALL SCALE OF THE STRAINS

In its application to precipitation hardening, the theory of Mott and Nabarro (1940) assumed that the regions in which the strain retained a roughly constant value were so large that an external stress could not cause a dislocation to move until it reversed the strain in each region through which the dislocation passed. Dislocations, pictured as straight lines since the early theories (Taylor, 1934; Orowan, 1934; Polanyi, 1934), must in fact be to some extent flexible, as has been emphasized by Burgers (1940). In accordance with the ideas of Mott (1946), the energy of a dislocation is an increasing function of its length, and this rate of change of energy with length leads to properties of the dislocation analogous to those of a soap film. A dislocation can assume a wavy form in which its radius of curvature is everywhere large compared with atomic dimensions under the influence of stresses small compared with the elastic constants of the material. In a precipitation-hardened alloy in which the stresses retain the same sign over distances of hundreds of atomic spacings, the equilibrium position of a dislocation is one in which each part of the dislocation lies in a region of the crystal where its elastic energy is low. To move such a dislocation involves lifting all parts simultaneously from potential troughs, and so requires an external stress $E\epsilon f$. Under small-scale stresses, such as are produced by solution hardening, the dislocation cannot follow the contours of minimum elastic energy, for this would require it to adopt a form in which its radius of curvature was of atomic dimensions, and so would require stresses of the order of the elastic constants. If the axis of a dislocation were a rigid straight line, the stress opposing its motion would vary in sign from region to region along its length and the resultant force would be small. For example, in a solid solution of atomic concentration f, the amplitude of the stress is of the order $E\epsilon f$, and this stress changes sign at the boundaries of regions having linear dimensions of the order of $f^{-\frac{1}{3}}$ atomic spacings. The expected mean value of such an oscillating function depends on the size of the region over which the mean is taken. The length of a dislocation is not known, but the fact that the yield stress of a single crystal is independent of the size of the specimen suggests that slip takes place independently in domains which are small compared with the dimensions of the specimen. The size of these domains is not known, but they may tentatively be identified with the mosaic elements of linear dimensions 10^{-4} or 10^{-5} cm. which are believed to form a real crystal. The stress required to move a straight dislocation 1000 atoms long is the mean of $1000 f^{\frac{1}{3}}$ contributions of random sign and of amplitude $E\epsilon f$, and so is of order $E\epsilon f/\sqrt{(1000 f^{\frac{1}{3}})}$. In a

5% solution, with $f=1/20$, this is about $1/20$ of $E\epsilon f$. It is possible that the mean stresses may be even smaller, for the solution in its state of lowest energy will tend to have foreign atoms as far from one another as possible, and so arranged at roughly equal distances apart. In the extreme case of an ordered solid solution, the components of stress along successive elements of the dislocation of individual length $f^{-\frac{1}{3}}$ lattice spacings cancel, and the resultant mean stress on a straight dislocation is of order only $E\epsilon f/1000f^{\frac{1}{3}}$, or $1/400$ of $E\epsilon f$. In many order-disorder transformations the disordered phase is cubic and the ordered phase is of lower symmetry. Within a single crystal, ordering starts from many nuclei, and the crystal in the ordered state is broken into domains differing in the orientation of their lattices. The resulting large-scale internal strains produce hardening effects comparable to those produced by precipitation (Barrett, 1943). However, in the case of β brass (Smith, 1942), in which the ordered lattice is cubic, annealing at room temperature produces a softening which has been attributed to the growth of the domains of coherent order.

According to these arguments, the elastic limit of a dilute (and therefore disordered) solid solution should be given by an expression of the form

$$S = \alpha E\epsilon f, \qquad \qquad \ldots\ldots(3)$$

where S, E, ϵ and f have their previous meanings, and α is a pure number equal to about 0·05, and almost independent of the concentration.

§4. COMPARISON OF THE THEORETICAL FORMULA WITH OBSERVATIONS

(i) *Au in Ag, Ag in Au.*

For this system the resolved shear stress across the glide plane in the glide direction is given by Sachs and Weerts (1930) in the form

$$\tau = 0·060k + 0·091(1-k) + 1·708k(1-k),$$

where τ is the stress in kg./mm², and k is the atomic concentration of Ag.

They also give the lattice parameter in the form

$$a = 4·07765k + 4·0700(1-k) - 0·240k(1-k).$$

Taking $E = 8 \times 10^{11}$ dynes/cm² for Au, $E = 7·7 \times 10^{11}$ for Ag, and substituting the values of $d\tau/dk$ and da/dk given by the above expressions when $k=0$ and when $k=1$ into equations (2) and (3) leads to $\alpha = 0·051$ for Ag dissolved in Au, $\alpha = 0·028$ for Au dissolved in Ag.

(ii) *Zn in Cu*

Here von Göler and Sachs (1929) give an elastic limit which initially rises linearly with the concentration of Zn, the slope of the stress-concentration curve being 0·1 kg./mm² for 1% addition of Zn. The lattice parameter rises linearly with concentration from $a = 3·6088$ A. for pure Cu to $a = 3·686$ A. for 67·5% Cu. These values give $\alpha = 0·015$.

(iii) *Zn in Al*

For this system Elam (1925) says that single crystals of Al + 18·6% Zn show a well-marked yield at 15 tons/sq. in. (which would correspond to a resolved stress of about 7½ tons/sq. in.). Accepting Elam's figures for the lattice parameters (Al 4·06 A., alloy 4·18 A.), and taking E for Al as 7×10^{11} dynes/cm² would lead

to $\alpha = 0.056$. However, a large expansion of the lattice on adding Zn to Al is unexpected, and is not supported by the results of other workers.   If Elam's figures for the lattice parameter are replaced by those of Owen and Iball (1934) (Al: 4·041 A., Al+19% Zn: '4·0345 A.), the value of α becomes 1·05.   Such a large value should be characteristic of precipitation hardening rather than solution hardening.   The solubility of Zn in Al is not known beyond dispute, but Hansen (1936) considers the most reliable results to be those of Schmid and Wassermann, leading to a solubility of less than 2% at room temperatures.   This gives considerable support to the suggestion that Elam's crystals were age-hardened.

(iv) *Cd in Zn*

In this system again the hardening found by Rosbaud and Schmid (1925) is remarkably great.   Their values for the elastic limit may be combined with the value $E = 8 \times 10^{11}$ dynes/cm² for Zn.   The value of ϵ may be estimated from the atomic volumes as $(13\cdot01 - 9\cdot16)/3 \times 9\cdot16$, and this leads to $\alpha = 0.22$.   This high value might be attributed to some peculiarity of the hexagonal lattice, but such an explanation seems unlikely, for in a single crystal specimen both hexagonal and face-centred cubic metals are known to slip on just one family of close-packed planes.   A more likely explanation is that the specimens were precipitation hardened, since Hansen accepts the estimate of Boas that the solubility of Cd in Zn is less that 0·4% at 150° c. and unmeasureably small at 100° c. or below. Rosbaud and Schmid assume that the Cd is in solution because no eutectic is visible in the microstructure.   On the other hand, the microphotographs of single crystals reproduced show large inclusions of some other phase or orientation. These may be twins (which would harden a single crystal of a hexagonal metal considerably), but are more likely to be laminar precipitates of another phase formed by diffusion.

(v) *Ni in Cu, Cu in Ni*

This system was investigated by Osswald (1933).   The curve relating elastic limit and concentration is distorted from the simple parabolic form which represents the results for the system Au–Ag.   Osswald finds a negligibly low elastic limit for copper, rising to a maximum resolved shear stress of 3·0 kg./mm² at a concentration of about 60 atomic % Ni, and falling to 0·6 kg./mm² for pure Ni. For dilute solutions, the slopes of the curves may be estimated as 0·2 kg./mm² for 1% Ni added to Cu, and 0·17 kg./mm² for 1% Cu added to Ni. The Young's moduli are 11×10^{11} dynes/cm² for Cu and 21×10^{11} dynes/cm₂ for Ni.   The lattice parameters are $a = 3\cdot608$ A. for Cu, $a = 3\cdot5175$ A. for Ni (Burgers and Basart, 1930; Owen and Pickup, 1934).   Taking the mean of the results of the authors quoted, Δa for 1% addition is -0.0013 A. for Ni added to Cu, 0.00078 A. for Cu added to Ni.   These values substituted in equations (2) and (3) give $\alpha = 0.049$ for Ni dissolved in Cu, and $\alpha = 0.036$ for Cu dissolved in Ni.

(vi) *Cu in Au, Au in Cu*

The elastic limits of polycrystalline alloys of copper and gold were determined by Broniewski and Wesolowski (1934).   1 atomic % Cu added to Au increases the elastic limit by 0·5 kg./mm²., and 1 atomic % Au added to Cu increases the elastic limit by 0·7 kg./mm²   The corresponding values of α are 0·073 and 0·066.

§5. APPLICATION TO HARDNESS MEASUREMENTS

A number of workers have measured the hardness of solid solutions (Goebel, 1922; Norbury, 1924; Frye and Hume-Rothery, 1942; Brick, Martin and Angier, 1943), and an interpretation of their results on the lines of the present theory would be of interest. However, in the Brinell test there is very heavy cold working of the metal close to the indentation: Norbury (1924) and Frye and Hume-Rothery (1942), by taking as their measure of hardness the " ultimate Meyer hardness ", measure a quantity which lies even farther from the initial elastic limit and closer to the limiting strength after heavy cold work than the Brinell number determined from a shallow impression. Attempts to extrapolate Frye and Hume-Rothery's results by Meyer's formula to give the hardness. corresponding to an impression of diameter 1/10 of that of the ball lead to very irregular figures, as was to be expected from the work of Norbury (1923) on annealed copper and copper alloys. It appears that the experimental results measure the limiting strength after heavy cold work rather than the elastic limit. In general, dilute solid solutions work-harden under small strains less rapidly than does the pure solvent. Elam (1927) finds that in single crystals of brass this is also true for large concentrations and large strains, but Brick, Martin and Angier (1943) find that rolling reductions of 20% or more produce increases in the hardness of polycrystalline solid solutions in copper which are linearly proportional to the excess of the hardness of the annealed alloy above that of pure copper.

The reason for the abnormally small work-hardening of solid solutions under small strains is not known. It is presumably connected with the fact observed by Masima and Sachs (1928) that, whereas in single crystals of pure metals the active glide plane gives way to a new plane as soon as the resolved stress on the new plane exceeds that on the active plane, in solid solutions the original plane remains active until the stresses on the new plane considerably exceed those on the original plane. Possibly the slip process itself produces a type of ordering of the solid solution which is conducive to slip on the active plane. An alternative possibility is that the initial irregularity of the lattice of the solid solution causes small areas of lattice planes to buckle under-small stresses, and so prevents the catastrophic rotations of large blocks of the lattice which lead to Laue asterisms in the early stages of slip in pure metals. It would be of interest to know if the early stages of slip in solid solutions, which are characterized by abnormally small hardenings, do in fact show abnormally little lattice distortion in Laue photographs.

Owing to this complicating factor, the analysis of hardness tests is not straightforward. It will be assumed that the Brinell or ultimate Meyer hardness exceeds. the elastic limit of a single crystal by a factor β, so that the hardness is given by

$$H = \alpha \beta E \epsilon f. \qquad \qquad \ldots \ldots (4)$$

The numerical value of β is not known with any accuracy, nor (except in special cases) is it certain that the value of β remains roughly constant for a series of solutions in the same solvent. A rough estimate of the magnitude of β may be based on the consideration that in general the Brinell hardness of a pure annealed metal is of order 40 kg./mm.² (e.g. Fe 67, Ag 30, Cu 45, Al 16), while the elastic limit for

a single crystal is of order 0.2 kg./mm² (e.g. Ag 0.06, Cu 0.1–0.4, Al 0.5), giving β about 200. Solid solutions showing little work-hardening would yield smaller values of β, and this estimate may be combined with one derived from a comparison of the measurements of Rosbaud and Schmid on the yield points of alloys of zinc and cadmium with the hardness measurements of Ludwik, which they quote. The two curves are not very similar in form, but the mean value of β is about 10. A figure of 80 will be taken as an estimate of the order of magnitude of β.

§6. SUMMARY OF THE EXPERIMENTAL OBSERVATIONS ON HARDNESS

Norbury (1924) examined the hardness of solid solutions of Ni, Zn, Al, Mn, Ag and Sn in copper, and similarly analysed the results of Goebel (1922) on solid solutions of Bi, Sn, Hg, and Cd in lead. In both cases the hardening effect is very roughly a linear function of the difference in size of the atoms of the solute and the solvent. In each series of experiments there was one anomaly, the solid solutions of Si in Cu and Na in Pb both being abnormally hard. Norbury attributes this to the formation of intermetallic compounds, so that the dissolved atom no longer has twelve symmetrically-placed nearest neighbours, but is closely bound to a smaller number of solvent atoms. This introduces stresses into the lattice which are more complicated than those which occur in other cases. Similar results have been obtained by Frye and Hume-Rothery (1942) on the hardness of solid solutions of Cd, In, Sn, Sb Mg, Al and Zn in silver. The solution of Al and Zn in silver decreases the lattice constant, whereas the other solutes increase the lattice constant. In all cases the hardness of solid solutions of various concentrations of these seven elements in silver is roughly a linear function of the change in lattice spacing (without regard to sign). Solutes which decrease the lattice spacing appear in general to produce a greater increase in hardness for a given change of lattice spacing than those which increase the lattice spacing. Frye and Hume-Rothery show that if the observations are grouped into series in which the dissolved atoms belong to the same row of the periodic table, and are present in the same concentration (2.4 atomic per cent or 5.0 atomic per cent), then the hardness in each series is very closely given by a linear function of the square of the change in lattice constant. The disagreement of the square law with the linear law originally suggested is discussed below. Finally, results by Brick, Martin and Angier (1943) for Si, Ni, Zn, Al, Mn, As, Sb and Sn dissolved in copper largely confirm the work of Norbury. Cobalt has an atomic volume slightly less than that of copper, and its solid solutions are very hard. The change in lattice parameter caused by adding Mg is not known, but the point for Mg falls on the curve relating hardening effect and difference of atomic volume.

A close connection between hardness and internal stress has recently been demonstrated by Frye, Caum and Treco (1944).

§7. EFFECT OF RESIDUAL STRESSES DUE TO OTHER CAUSES

The theory originally proposed led to a hardness of the alloy directly proportional to the deformation of the lattice, whereas the measurements of Frye and Hume-Rothery show a linear relation between hardness and the square of the deformation. These are in fact both limiting cases of the same general formula, and are obtained when the increase in hardness caused by the presence of the dissolved atoms is very large or very small in comparison with the hardness of the pure solvent.

Whether the dislocation is imagined as rigid or as flexible, the stress required to cause slip in a metal by moving an appreciable number of dislocations in it is determined less by the extreme values of the internal stress than by the average fluctuations of stress in the material. These are conveniently measured by the mean square stress, since the mean stress over the whole body must be zero.

Consider a pure metal containing random internal stresses (due to misfits at grain boundaries or imperfect annealing) of mean square value S_0^2. The hardness of this metal is given by

$$H_0 = \alpha\beta S_0, \qquad \qquad \dots\dots(5)$$

where α is a pure number, the value of which, according to the previous arguments, must be about 1 if dislocations are flexible and about $1/20$ or less if dislocations are rigid, and β is the ratio of the hardness to the elastic limit, and is of order of magnitude 50 to 100. If foreign atoms are now dissolved in the metal they give rise to a further random distribution of stress of mean square value $(E\epsilon f)^2$. The hardness of the solid solution is now given by

$$H = \alpha\beta\sqrt{(S_0^2 + E^2\epsilon^2 f^2)}. \qquad \qquad \dots\dots(6)$$

Eliminating S_0 between (5) and (6) gives

$$H^2 = H_0^2 + (\alpha\beta E\epsilon F)^2. \qquad \qquad \dots\dots(7)$$

The original theory, which neglected H_0, made H a linear function of $E\epsilon f$, and Goebel's results for solutions in lead, where the initial hardness is small, show a roughly linear dependence of hardness both on concentration (except where the solute is Hg) and on difference of atomic volume (except in the case of Na, which has already been attributed to the formation of a chemical bond).* Norbury's solid solutions in copper and Frye and Hume-Rothery's solutions in silver showed only small increases in hardness above that of the pure solvent. For such cases equation (6) may be written approximately as

$$H - H_0 = (\alpha\beta E\epsilon f)^2 / 2H_0. \qquad \qquad \dots\dots(8)$$

Norbury's results show an increase in hardness which for dilute solutions of a given element varies more rapidly than a linear function of the concentration in the cases of solutions of Si, Al, Ni and Zn. The effect is not shown in solutions of Sn, Ag, and Mn, but in the cases of Sn and Mn he states that " there are reasons for thinking that . . . . the results really lie on a curve bending upwards initially ", and in the case of Ag " it is possible that . . . . the results lie on a curve bending upwards and not on a straight line ". The results of Brick, Martin and Angier on 1% solutions in copper are well represented (except for Si) as a function of ϵ by a formula of type (7), and finally the quadratic law of Frye and Hume-Rothery is in agreement with the form of equation (8).

However, the dependence of hardness on concentration in the results of Brick, Martin and Angier for solutions in copper is linear even for small concentrations which increase the hardness by an amount less than the initial hardness of pure copper, and it seems in general that the hardening effects of solution and of other causes combine sometimes linearly and sometimes as independent random factors. It is to be expected that the small-scale strains due to solution will combine in a random manner with other small-scale strains, but will add linearly to the hardness produced by large-scale strains.

* Kurnakow *et al.* (1940) have confirmed that small additions of Na harden lead appreciably, while altering the lattice parameter very little.

§8. ANALYSIS OF EXPERIMENTAL HARDNESS MEASUREMENTS

As a check on the order of magnitude of the effect, the quantity $\alpha\beta$ of equation (8) may be estimated from the results of Frye and Hume-Rothery. They give $H_0 = 43$ kg./mm².$= 4\cdot2 \times 10^9$ dynes/cm.² and E for silver is $7\cdot7 \times 10^{11}$ dynes/cm.² The lattice spacing a in silver is $4\cdot1$ A., and so $\epsilon f = \Delta a/4\cdot1$. For $f = 5\%$. Frye and Hume-Rothery give two lines, one of slope $\Delta H/(\Delta a)^2 = 2\cdot4 \times 10^5$ kg./mm².$/$A.², the other of slope $0\cdot45 \times 10^5$ kg./mm².$/$A². (The factor 10^5 appears in their paper as 10^7, but this is an error in computation.) These correspond respectively to $\Delta H/\epsilon^2 f^2 = 4 \times 10^{14}$ dynes/cm.² and $0\cdot74 \times 10^{14}$ dynes/cm.² So the value of $\alpha\beta = (2H_0\Delta H/\epsilon^2 f^2)^{\frac{1}{2}}/E$ is $2\cdot4$ or $1\cdot0$.

It is also possible to estimate $\alpha\beta$ from Goebel's results on solutions in lead, which give hardening as a linear function of the distortion of the lattice. Norbury's plot of Goebel's results is expressed in arbitrary units, which appear to be hardening h in tenths of a kilogram per square millimetre and difference of size of solvent and solute atoms e in cubic centimetres per gram atom. This gives $H = 0\cdot981 \times 10^7 h$ dynes/cm²., $\epsilon = e/3 \times 18\cdot27$. For $f = 0\cdot01$, the points plotted by Norbury show $h/e = 2\cdot7$, corresponding to $H/\epsilon = 1\cdot45 \times 10^9$ dynes/cm.² Taking E for lead as $1\cdot6 \times 10^{11}$ dynes/cm.² gives $\alpha\beta = H/E\epsilon f = 0\cdot9$. This value of $\alpha\beta$ is consistent with that deduced from the results of Frye and Hume-Rothery. Combining these results with the rough estimate $\beta = 80$ gives values of α of the order of $0\cdot02$, in agreement with the estimates from the yield points of single crystals.

ACKNOWLEDGMENT

The writer is indebted to Professor N. F. Mott, F.R.S. for many stimulating discussions on this problem.

REFERENCES

BARRETT, C. S., 1943. *Structure of Metals* (McGraw-Hill).
BRICK, R. M., MARTIN, D. L. and ANGIER, R. P., 1943. *Trans. Amer. Soc. Met.*, **31**, 675.
BRONIEWSKI, W. and WESOLOWSKI, K., 1934. *C.R. Acad. Sci., Paris*, **198**, 569.
BURGERS, J. M., 1940. *Proc. Phys. Soc.*, **52**, 23.
BURGERS, W. G. and BASART, J. C. M., 1930. *Z. Kristallogr.*, **75**, 155.
ELAM, C. F., 1925. *Proc. Roy. Soc.*, A, **109**, 143; 1927. *Ibid.*, **115**, 148.
FRYE, J. H., CAUM, J. W. and TRECO, R. M., 1943. *Trans. Amer. Inst. Min. Met. Engrs.*, **152**, 83.
FRYE, J. H. and HUME-ROTHERY, W., 1942. *Proc. Roy. Soc.*, A, **181**, 1.
GOEBEL, J., 1922. *Z. Metallk.*, **14**, 357, 388, 425, 449.
GÖELER, FRHR. V. and SACHS, G., 1929. *Z. Phys.*, **55**, 581.
HANSEN, M., 1936. *Aufbau der Zweistofflegierungen* (Berlin : Springer).
KURNAKOW, N. S. *et al.*, 1940. *Ann. Secteur Anal. Phys.-Chim. U.S S.R.*, **13**, 233
MASIMA, M. and SACHS G., 1928. *Z. Phys.*, **50**, 161.
MOTT, N. F., 1946. *J. Inst. Metals*, **72**, 367 (May Lecture).
MOTT, N. F. and NABARRO, F. R. N., 1940. *Proc. Phys. Soc.*, **52**, 8.
NORBURY, A. L., 1923. *J. Inst. Metals*, **29**, 407, 423.
NORBURY, A. L., 1924. *Trans. Faraday Soc.*, **19**, 586.
OROWAN, E., 1934. *Z. Phys.*, **89**, 605, 614, 634.
OSSWALD, E., 1933. *Z. Phys.*, **83**, 55.
OWEN, E. A. and IBALL, J., 1934. *Phil. Mag.*, **17**, 433.
OWEN, E. A. and PICKUP, L., 1934. *Z. Kristallogr.*, **88**, 116.
POLANYI, M., 1934. *Z. Phys.*, **89**, 660.
ROSBAUD, P. and SCHMID, E., 1925. *Z. Phys.*, **32**, 197.
ROSENHAIN, W., 1921. *Proc. Roy. Soc.*, A, **99**, 196.
SACHS, G. and WEERTS, J., 1930. *Z. Phys.*, **62**, 473.
SMITH, C. S., 1942. *Metals Technol.*, T.P. 1517.
TAYLOR, G. I., 1934. *Proc. Roy. Soc.*, A, **145**, 362.

ON METHODS FOR THE CONSTRUCTION OF NETWORKS DUAL TO NON-PLANAR NETWORKS

By A. BLOCH

(Communication from the Staff of the Research Laboratories of
The General Electric Company Limited, Wembley, England)

ABSTRACT. Two networks are said to be dual to each other if the following correspondence exists between them :—

(a) Each element of one network has a counterpart in the other one.

(b) The currents through the elements in one network and the voltages across their counterparts in the other network are proportional to each other.

A network dual to a given one is usually constructed by connecting elements of the appropriate kind in a circuit the diagram of which forms the topological counterpart to the diagram of the original circuit. If this original circuit is a " non-planar " one (i.e. if it cannot be drawn without crossings between some of its branches), then there exists no topological dual, and for this reason it has been said that there exists no dual to a non-planar electrical network. However, if we are interested in an electrical network not because its circuit diagram has certain geometrical properties, but because it constitutes an assembly of a number of passive and active elements which perform in a certain manner (on account of their impedance properties and of certain relations of mutual constraint imposed by Kirchhoff's mesh and junction relations), then there is really no restriction of this kind.

We need only convert the original network into another with the same elements and the same performance, but with a planar circuit diagram and with additional constraints implemented by ideal transformers. This equivalent network can be then converted into its dual counterpart, which is another planar network subject to the dual counterparts of the transformer constraints of the first case. In this way we arrive at an assembly which performs in the required dual manner.

Introducing thus the concept of physical duality as opposed to geometrical duality, the paper explains in an elementary manner a number of ways in which this conversion into equivalent planar networks can be achieved. These methods are :—

(1) A method originally published by Julia.

(2) The dual counterpart of this method.

(3) The *method of fictitious junction points* (an electrical analogue to F. Schur's method of fictitious junction points in pin-jointed frameworks).

(4) The *method of the transferred terminal.*

(5) The general *method of separation.*

Some of these methods require a smaller number of ideal transformers than Julia's original method.

The paper ends with a short discussion of the errors introduced when the dual network is set up on a network analyser and when the ideal transformers have to be replaced by real ones ; the numerous methods available for the realization of an equivalent network enable a choice to be made so as to minimize these errors, or possibly eliminate them altogether.

The existence of a dual to a non-planar network is of some interest in connection with electro-mechanical analogies as it removes a restriction to which the " direct " analogy would otherwise be subjected.

§1. INTRODUCTION: THE PRINCIPLE OF DUALITY

TWO networks are said to be dual to each other if the following correspondence exists between them :—

(*a*) Each element of one network has a counterpart in the other.

(*b*) The currents through the elements of one network and the voltages across their counterparts in the other network are proportional to each other.

In other words, we have the following relations :—

$$i_\kappa' \, R_0{}^1 = e_\kappa'' \, , \qquad \qquad \ldots \ldots (1)$$

$$i_\kappa'' R_1{}^0 = e_\kappa' \, , \qquad \qquad \ldots \ldots (2)$$

where i_κ' and e_κ' are current and voltage at element κ' of the first network and i_κ'' and e_κ'' current and voltage at the corresponding element κ'' of the second network, and where $R_0{}^1$ and $R_1{}^0$ are two constants (each of the dimension of a resistance) which are characteristic of the particular correspondence under consideration.

It is obvious then that the behaviour of one network is a copy of that of the other network, but with the rôles of current and voltage interchanged.

We shall see in the following that the existence of such a relation can be predicted by very simple criteria, and for this reason it will obviously be useful, for it will enable us to predict the performance of one network without a preceding analysis provided we know the behaviour of its dual counterpart. This *Principle of Duality* has been known for some time. The phrase seems to have been introduced into electric network theory by H. Sire de Vilar (1901). The first textbook dealing with it was A. Russell's *Theory of Alternating Currents* (1904), which gives a large number of applications and a list of earlier references. There are, or rather were, however, certain difficulties in the application of the principle whenever one of the networks concerned was " non-planar."*

These difficulties give rise to an interesting question in connection with electro-mechanical analogies (Bloch, 1945). There are two types of such an analogy: one in which we represent a force by a voltage (the " direct " analogy), and one in which we represent a force by a current (the " inverse " analogy). If we represent a given mechanical system once by the first and once by the second type of analogy then we obtain two electrical networks which are—obviously—in dual relationship to each other. Now the circuit diagram of the inverse type of analogy can always be obtained as a simple translation of the mechanical layout if this is drawn according to certain conventions. Should this circuit diagram be of the non-planar type, and were we to take the statement literally that there exists no dual to such a network,† then we would have to acknowledge that the field of application of the direct analogy was restricted. We shall see that this is not the case, and that the apparent contradiction turns on the definition of the concept of duality. Before concentrating on the special

---

* A non-planar network is one the circuit diagram of which cannot be drawn without crossings between some of its branches.

† Whitney (1931) shows that there exists no geometrical dual to a non-planar graph; Barnes and Gardener (1932) seem to take this as the basis of their statement that there exists no dual to a non-planar network.

problem of the present paper we shall therefore deal first more fully with the normal and general aspects of the duality principle.

It will be seen that we have to distinguish between duality in a topological sense and duality in a physical sense.   By duality in a physical sense we have in mind a correspondence between two assemblies of elements such as outlined above, a correspondence which is concerned only with the behaviour of the elements specified.   By duality in the topological sense we have in mind a relation between the structure of the two circuit diagrams such that each branch of one network has its counterpart in the other network and that all those branches of one network which radiate from a particular junction point have counterparts in the other network which form a closed mesh.   The fact that the branches of the network are linked together in a certain way imposes on the currents therein and on the voltages across them certain conditions of constraint commonly known as Kirchhoff's equations.   These conditions of constraint are obviously decisive for the behaviour of the network.   Now we shall see that in order to bring about dual behaviour of two networks it is sufficient to establish a certain impedance relation between corresponding elements of the two networks, to provide external sources of dual character, and finally to provide conditions of constraint which are of dual character.   If we insist in bringing about this dual type of constraint by an appropriate structure of the circuit diagram—and by this means only—then we shall see that the concepts of physical and of topological duality are inseparably linked.   On the other hand, if we are only interested in the behaviour of the networks (i.e., assemblies of elements), such topological duality is immaterial; we may then introduce the required conditions of constraint by other means.   These other means are ideal transformers, and we shall see by their use it will always be possible to specify one or more networks which are dual in the physical sense.   Thus the theorem that there exists no topological dual to a non-planar network does not occupy the dominant position which one might at first assume.   It is nevertheless still of interest, for obviously topological duality is the simplest way of bringing about dual conditions of constraint, and we shall always try first to achieve it.   Knowledge of the theorem will, however, save us from useless attempts to find this ideal solution.

The present paper discusses first R. Julia's method for deriving a physically dual network by the introduction of ideal transformers.   By giving the justification for this procedure more fully than it was given in his original publication we are automatically led to further methods of achieving the same result.   This is useful, for some of the new methods require a smaller number of ideal transformers than Julia's original method.   At the same time the fact that we are thus able to choose between a whole series of different embodiments of the dual network will be useful when we attempt to set up the dual network in a network analyser, for some of these alternative ways will possibly be less susceptible to the imperfections introduced by the use of real transformers in place of the ideal ones.

§2. CRITERIA OF DUAL CORRESPONDENCE: CONSTRUCTION OF THE DUAL COUNTERPART IN THE CASE OF PLANAR NETWORKS

The existence of the dual relationship between two networks as specified by equations (1) and (2) can be predicted very simply if we pause for a moment

and consider what data determine the behaviour of a given network. There are:—

    (a) The relations between the currents through and the voltages across the various elements of the network, i.e. a series of equations of the type

$$i_\kappa' = f(e_\kappa'); \qquad \ldots\ldots(3)$$

    (b) The way in which these elements are connected, i.e. the relations of constraint which exist between the currents meeting at a given junction point and the voltages which we encounter when we pass around a closed mesh (Kirchhoff's equations);

    (c) The initial state of the network (currents through and voltages across certain elements);

    (d) The external " causes " (voltages or currents) impressed on the network.

All these data together give us a system of equations—not necessarily a linear one—the solution of which describes the behaviour of the network.

Suppose now that we have a second network for which we can list a corresponding set of data, the correspondence being of the type described by equations (1) and (2). This means that we get a corresponding set of data which is of the same form as the original one but with the difference that the symbols i and e have changed places. It is obvious then that the resultant system of equations will show the same interchange of symbols, and the same holds for the final solution of this system. Hence the second network must show a behaviour dual to that of the first network.

Let us now discuss in detail what is involved in having such an interchanged set of data between two networks. In order not to interrupt the exposition we shall deal first with the items (c) and (d). The dual counterpart of the initial conditions is self-evident; any statement of this type can be translated in a straightforward manner by the application of equations (1) and (2). The same applies to the specification of the external " sources " which are connected to the network. A voltage source corresponds to a current source of proportional strength and vice versa.

As regards (a), equation (3) is, according to equations (1) and (2), replaced by an equation of the type

$$\frac{e_\kappa''}{R_0^1} = f(i_\kappa'' R_1^0), \qquad \ldots\ldots(4)$$

i.e., apart from factors of proportionality, the graph of current against voltage for each element of the original system will represent the graph of volts against current in the new system. If we limit ourselves to linear systems and study the steady-state behaviour of such systems at a frequency ω we have for any particular element of the first system and for its counterpart the relations

$$e_\kappa' = i_\kappa' Z_\kappa' \quad \text{and} \quad e_\kappa'' = i_\kappa'' Z_\kappa'', \qquad \ldots\ldots(5)$$

where Z_κ' and Z_κ'' denote the impedances of the elements concerned. Multiplying these two equations and taking into account equations (1) and (2), we obtain

$$R_0^1 R_1^0 = Z_\kappa' Z_\kappa''. \qquad \ldots\ldots(6)$$

Equivalent expressions for the last equation are

$$Y_\kappa' = Z_\kappa'' \frac{1}{r_0^2},$$

$$Y_\kappa'' = Z_\kappa' \frac{1}{r_0^2},$$

$$\left.\right\} \qquad \ldots\ldots(7)$$

where $\qquad Y_\kappa' = 1/Z_\kappa' \quad$ and $\quad Y_\kappa'' = 1/Z_\kappa'',$

and where $\qquad r_0^2 = R_0' R_1^0.$ $\qquad \ldots\ldots(8)$

From these equations it follows that the dual counterpart of a resistance R is a conductance $G = R/r_0^2$, that the counterpart of an inductance L is a capacitance $C = L/r_0^2$ and that of a capacitance C an inductance $L = Cr_0^2$.

It will be noted that these relations are independent of the frequency ω, i.e., that dual relationship once found for certain frequencies will be valid for all frequencies. On account of Fourier's theorem this means that once the dual relationship is established for steady-state conditions at one particular frequency it will hold good not only for steady-state operation at any other frequency but also for transient behaviour.

Turning now to point (*b*), the problem of interconnection of elements, we will discuss a special case first, namely the case where two or more elements in the first system are connected in series. This means that all these elements carry the same current, and that their voltages add. Their dual counterparts must then be so connected that across each of them appears the same voltage, and that all their currents add, i.e. they must be connected in parallel. This relation works, of course, the other way round. It is a very useful relation, for it can be applied for checking even when we have to deal with a more general type of network; we can always give to some of the impedances of a network the values zero or infinity, and thereby derive from the given network a number of simpler networks to which these conversion rules can be applied. (We might establish here a principle of " detailed reciprocation ", the name being chosen in analogy to the " principle of detailed balancing " as used in thermodynamics).

A more general way of establishing relations of constraint dual to the given ones is based on a consideration of Kirchhoff's equations. Suppose we have for a junction point (σ) of the given network written down the condition that the sum of all branch currents reaching this point is zero, i.e.

$$i_{\sigma 1}'' + i_{\sigma 2}'' + \ldots = 0. \qquad \ldots\ldots(9)$$

It follows then that in the dual network the voltages across the counterparts of all these branches must add up to zero:

$$e_{\sigma 1}'' + e_{\sigma 2}'' + \ldots = 0, \qquad \ldots\ldots(10)$$

i.e., we would normally expect these counterparts to be so arranged that they formed a closed mesh. Conversely we see that elements which are so arranged that they form a closed mesh should have dual counterparts which are joined to a common junction. We see that dual conditions of constraint are automatically introduced if the second network is in this particular geometrical relation to the first one, i.e. if it is the topological dual of the first network.

There exists a routine method for drawing such a topological dual (Cauer, 1934).

In the interior of each mesh of the given network we select a point. These points will form the junction points of the new network; we need only join each of these points to its neighbours (i.e. to those new points situated in adjoining meshes) in order to complete the layout of the dual network. Each new branch crosses one old branch (the branch which was common to the two adjoining meshes). The procedure is illustrated in figures 1*a* and 1*b*. In figure 1*a* the new circuit is sketched in dotted lines, in figure 1*b* it is redrawn, and at the same time the proper elements inserted, taking into account the relations (6) and (7) and their consequences as to the interchange of capacities and inductances. It is seen that in a case where the original branch is composed of several elements in
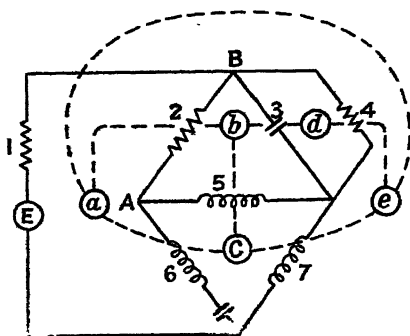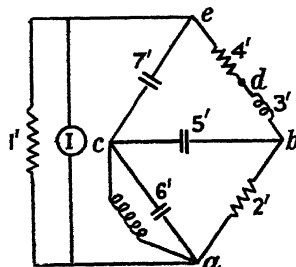
Figure 1 *a*.   Figure 1 *b*.

series (e.g. 4 and 6) we could draw such a crossing line through each of these elements; all these new branches are then in parallel, and if we place into each of these new branches an element with an admittance proportional to the impedance of the "crossed" old element the resulting compound admittance will obviously be the correct dual of the original compound impedance.*

The procedure just described breaks down if the original circuit diagram cannot be drawn without crossings between some of the lines (" non-planar " networks). Such networks have no topological duals, and in their case we have to use other means for introducing the required constraints between the various elements of the dual network.

§ 3. IDEAL TRANSFORMERS

We shall now enter into a discussion of these other means which will enable us to introduce the required constraints. Suppose we have an iron core with several windings of the same number of turns, i.e. a multi-winding transformer, schematically represented in figure 2 *a*. Let us assume further that the permeability of the core material is infinite and that the windings have negligible leakage. In this case we are then assured that the same flux is linked with each of the windings, i.e. that the terminal voltages of all the windings are equal. Furthermore, the finite flux required for any such finite voltage can then be

---

* It will be noted that this method of drawing the dual circuit can also be used to define in a consistent manner the sense of the dual currents or voltages. We need only agree that the arrows defining this sense in the new branches must be produced from those of the original network by, say, a rotation in a clockwise rotation.

produced by a magnetomotive force which is infinitesimally small, i.e. we must have

$$i_1 + i_2 + i_3 + \ldots = 0. \qquad \ldots\ldots(11)$$

A transformer of this kind is called an *ideal transformer*. Real transformers differ from the ideal one by possessing leakage inductance which we might visualize as a separate inductance connected in series with each individual winding and by requiring a finite magnetomotive force for the creation of their magnetic flux. With real transformers the voltages across the individual windings may then no longer be exactly equal, and furthermore, equation (11) will be only approximately fulfilled. These deviations are, however, usually small, and they can possibly be allowed for if it is desired actually to build the dual network. For a discussion of principles we are certainly allowed to make use of the concept of an ideal transformer, and we see that a transformer of the type just described imposes on the currents through its windings the same constraints as if they were associated with a common junction point.

The dual counterpart to such a transformer is shown in figure 2 *b*. There the magnetic circuit is so arranged that the cores of the individual coils—which
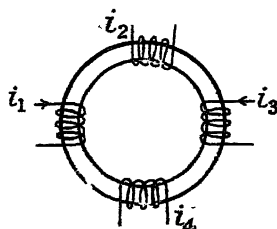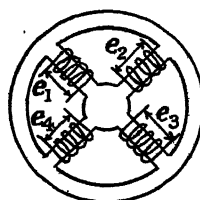
Figure 2 *a*.   Figure 2 *b*.

in figure 2 *a* were all connected in series—are all connected in parallel. In this magnetic circuit, the sum of all the magnetic fluxes which arrives at one of the magnetic junctions must be zero, and from this it follows that the sum of all the voltages of the windings of the various cores (which individually are proportional to the fluxes of the corresponding cores) must be zero:

$$e_1 + e_2 + e_3 + \ldots = 0. \qquad \ldots\ldots(12)$$

Associated with this condition is the condition that the currents through all the windings must be equal; otherwise the magnetomotive force of the individual limbs would not be equal and—in view of the assumed infinite permeability of the core material—the difference in the m.m.f. of any two of the limbs would drive a magnetic flux of infinite magnitude around the closed circuit formed by these two limbs. A transformer of the second type enforces, therefore, the same type of constraint as results from the existence of a closed mesh. Transformers with two windings are obviously special cases of these general multi-winding transformers, and we see that in their case the two types are identical.

Transformers with two windings are easier to build so as to approach the requirements of zero leakage inductance. It is therefore of interest—for those cases where it is intended to set up a physical circuit—that the same results that can be obtained by a multi-winding transformer with *n* windings can be obtained

by the use of $n-1$ transformers with two windings only. If we start with n transformers with two windings we need only connect all the secondaries in series forming a closed loop, in order to enforce on all the primary voltages the condition that their sum is zero. On the other hand, if we connect all the secondaries in parallel we enforce the condition that the sum of all the primary currents is zero. As figure 3 shows (for the case $n=3$), one of those transformers is actually redundant. (For the purpose of derivation it is often convenient to retain n transformers and to carry out the reduction to $n-1$ as a last step.)

Having thus shown that these transformers will enable us to enforce constraints of the " mesh " and " junction " type,* we have to consider how the method, described in the preceding section, for the routine derivation of the dual network can be operated in the presence of such circuit elements. The dual network we would expect to be subject to transformer constraints which are the dual counterparts of the transformer constraints of the original network. It will be seen that this routine method automatically takes care of the situation provided we stipulate that each transformer coil encountered in this process is taken to
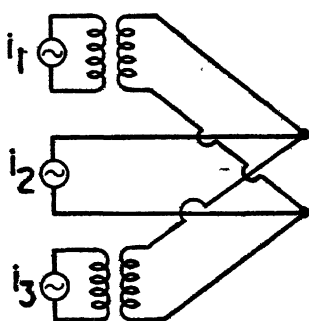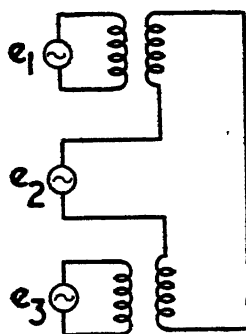
Figure 3 *a*.   Figure 3 *b*.

have as counterpart another transformer coil assembled on an ideal transformer of the dual type. Suppose we have in the first network a number of elements $Z_1', Z_2', Z_3', \ldots$, each with such a transformer coil in series, the transformer itself being of the junction type, which insures that the sum of the currents $i_1'', i_2'', i_3'', \ldots$ through all these elements is zero. In the new network these coils will automatically be placed in parallel to the dual counterparts $Y_1'' \ Y_2'' \ Y_3''$ of the elements $Z_1', Z_2', Z_3' \ldots$, and there by enforce the condition that in the dual network the voltages $e_1'' \ e_2'' \ e_3'' \ldots$ across these elements shall add up to zero. Obviously the method would have worked just as well the other way round, and would have converted a constraint of the mesh type into one of the junction type.

It is perhaps of interest to note that the change over from the one type of transformer to its dual counterpart is actually included in the routine process of " dualization " provided we pay attention not only to the electrical but also to the magnetic circuits, the junction type of transformer having a magnetic circuit dual to that of a mesh type (figures 2 *a* and 2 *b*). If we replace the multi-winding transformer by two-winding transformers the exchange of types is simply a consequence of " dualizing " the connection of the secondary windings.

* We shall for this reason refer to such transformers as mesh or junction-type transformers.

§4. THE CASE OF NON-PLANAR NETWORKS

(a) *The general principle*

We are now in a position to deal with non-planar networks. Networks of this type can be converted in various ways—presently to be described—into equivalent planar networks which are subject to additional constraints by the aid of one or several ideal transformers. By *equivalence* we mean here that the resultant networks possess, apart from those specially added ideal transformers, the same elements as the original one, and that the currents and voltages at these elements are identical with those of the original network. This equivalent network is then converted into its dual counterpart by the routine method already described.

Of the various methods to convert a non-planar network into an equivalent planar network we shall discuss here the following—without making an attempt to be exhaustive: Julia's method (1939); a method dual to Julia's method; the method of fictitious junction points; the method of the transferred terminal; and the general method of separation.

(b) *Julia's method*

This method is characterized by the following steps:—

(1) Select a suitable junction poi it and introduce into each of the branches. leading to this point the coil of a junction-type transformer. This will not alter the current distribution in the network as the constraint exercised by this transformer is identical with that already in existence. In figure 4 *b* the junction point selected was point A of the original network, illustrated in figure 4 *a*.

(2) Short circuit one of the branches leading to the junction point selected. This has been indicated in figure 4 *b* by the dotted line which connects A to D, and short circuits branch 1. This step again will not alter the current distribution of the network; though we have lifted by this step the constraint originally exercised by the existence of the junction point A, the transformer coils exercise now an identical constraint. Since the sum of the currents flowing through the ideal transformer windings is zero, there can be no current through the short-circuiting link.

(3) Duplicate the short-circuiting link in the way shown in figure 4 *c*, introducing thereby the auxiliary junction points A' and D' (figure 4 *c* shows only those parts of the original network which are essential to the argument).

(4) Cut the link AA'; this cannot alter the current distribution in the network, because the potential difference AA' after the cut is still zero on account of the connection DD'. But now we are allowed to cut DD' as well, for the current through this isthmus must, of course, always be zero.

By these steps we have split the original network into two entirely separate networks, the elements of which have still the same voltages across and the same currents through them as they had before. (We need not have cut DD', for after the cut AA' has been made it would have been possible simply to fold aside the loop containing Z_1 and produce thereby a plane configuration.)

The new plane configuration is redrawn in figure 4 *d*. We may check that on account of the transformer action the current through element Z_1 is still

equal to the algebraic sum of the currents through elements Z_2 and Z_3; further-more, that the voltage drop in going from D to B is equal to the voltages across elements Z_3 and Z_1, and similarly the voltage drop in going from D to F is equal to drops across Z_2 and Z_1.
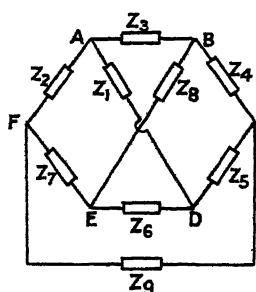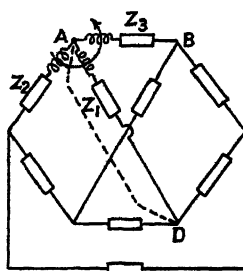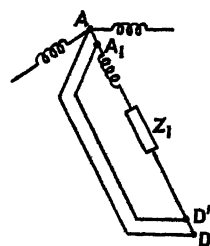
Figure 4 a.   Figure 4 b.   Figure 4 c.
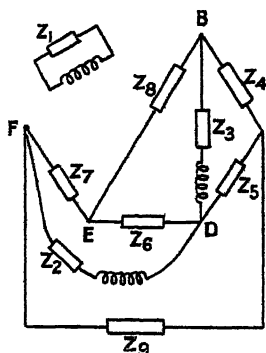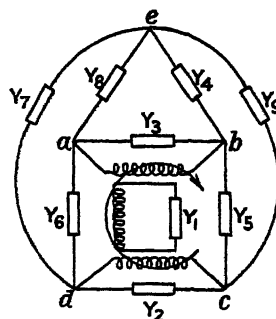
Figure 4 d.   Figure 4 e.
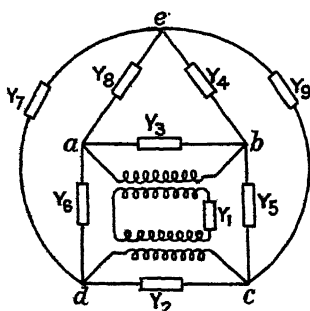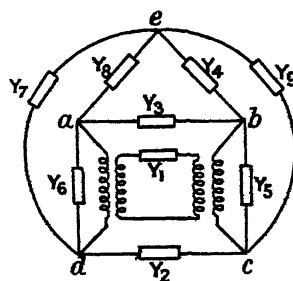
Figure 4 f.   Figure 4 g.

Now the new configuration of elements, being a planar one, can be trans-formed into its dual counterpart by the method described above. We arrive then at the network shown in figure 4 e. Figure 4 f shows the same network with two 2-winding transformers. Figure 4 g is included for the sake of interest: it shows the final network at which we would have arrived had we eliminated from the original network the junction point D instead of A.

(c) *The dual counterpart to Julia's method*

We come now to the description of a new method for the derivation of an equivalent planar network which can be characterized as the dual counterpart of Julia's method. In Julia's method we used transformer action to replace a junction type of constraint. Consequently, the dual network arrived at is one in which transformer action is used to implement a mesh type of constraint.

Obviously it is possible to invert this procedure, i.e. we can start by replacing a mesh type of constraint by transformer action and finish them with a dual network in which transformers implement a junction type of constraint. To
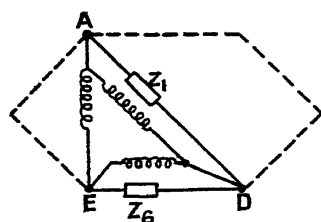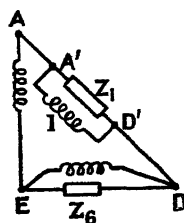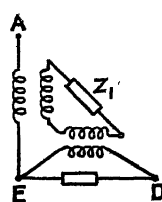
Figure 5 a.   Figure 5 b.   Figure 5 c.
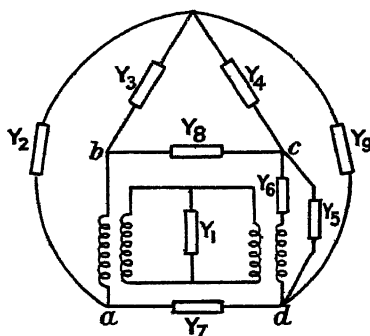
Figure 5 d.

illustrate this procedure we shall carry it out with the same network which we used just now, and we shall again eliminate from this network branch 1. To achieve this we proceed in the following steps:—

(1) Select a suitable mesh which contains the branch to be eliminated. The mesh to be selected may also include such branches of the original network as contain elements of infinite impedance. Thus, for instance, we may select in figure 4 a the mesh ADEA, of which the branch AE is of infinite impedance, and therefore not explicitly shown as " branch " in the diagram. Place in parallel to each mesh branch the coil of a mesh type of ideal transformer (or the input coil of an ideal two-winding transformer, the secondaries of which are all connected in series). This step, illustrated in figure 5 a,* cannot upset the current distribution in the network, for the constraint thereby imposed is identical with one already in existence.

(2) Re-draw this network in the way shown in figure 5 b, introducing thereby the auxiliary junction points A' and D'. Cut the link AA'. This further step

_____

* The figure shows only those branches which are relevant to the argument.

is permissible without upsetting the network performance as, on account of the transformer action, the potential difference across coil 1 is maintained independent of any current flowing therein; hence the potential difference AA' after the cut must still be zero.*

(3) Cut the link DD'; this is permissible now—after the cut AA' has been carried out as the current in the isthmus DD' must be zero.

By these steps we have again removed branch 1 from the original network. Figure 5 c shows how the same end can be achieved by the use of 2 two-winding transformers. Here it is immediately obvious that the two transformers supply a potential AE + ED = AD to element 1, and that the current drawn by this element is in fact still drawn (via the primaries) from junction point A and returned to junction point D.

Figure 5 d shows the dual to this latter network. There is no need to illustrate separately the dual to the network shown in figure 5 b.

(d) *The method of the fictitious junction point*

This method can be considered as a generalization of Julia's method; it will as a rule lead to a smaller number of ideal transformers.

Julia's method was characterized by the short circuiting of a branch—this short circuit being without influence on the current distribution on account of the transformer constraint previously introduced. The new method introduces a short circuit at the cross-over of two or more suitably selected branches. By converting thus all the cross-overs in the network into real junction points, the network ceases to be a non-planar one. The original current distribution is again maintained by the introduction of suitable transformer constraints.

The principle of the method is illustrated in figures 6 a to 6 c. Figure 6 a shows two branches of the original network which cross each other without making contact. In figure 6 b this cross-over has been converted into a real junction point (X); the original current distribution has nevertheless been maintained by the insertion of the primary and secondary coil of an ideal transformer into the lead BD. The action of this transformer ensures that as much current leaves the new junction point X through Z_4 as arrives there through Z_2. Hence none of this current can leave through Z_1 or Z_3. (The introduction of another transformer into the branches containing the elements Z_1 and Z_3 would of course be permissible, and may in a practical case aid in the maintenance of this state of affairs). Figure 6 c finally shows the dual counterpart of the network section shown in figure 6 b.

An application of the principle just discussed to the network of figure 4 a is shown in figures 7 a and 7 b. It will be noted that by this method we save—compared with the solutions previously given—one ideal transformer. This saving in a number of ideal transformers is larger still in the case of the network shown in figure 8 a which represents the second example given in Julia's paper. In dealing with this network Julia removes the junction point A, and requires

---

* A somewhat more detailed argument is as follows: As the potential across coil 1 remains unchanged, the current through element Z_1 remains also unchanged. After the cut, this current must be supplied from the transformer coil 1, and for this reason all the other coils must then carry a current equal to this current. This means that at junction point A a current equal to this current is still leaving (into coil AE), and at junction point D a current of equal magnitude is still arriving.

for this purpose either an ideal transformer with four windings or three transformers with two windings. From inspection of figure 8 a it follows immediately

Figure 6 a.

Figure 6 b.

Figure 6 c.

Figure 7 a.

Figure 7 b.

Figure 8 a.
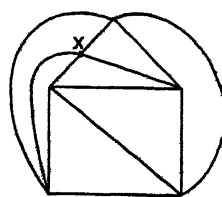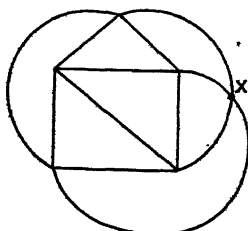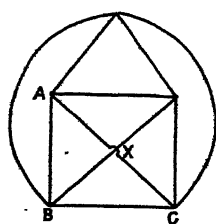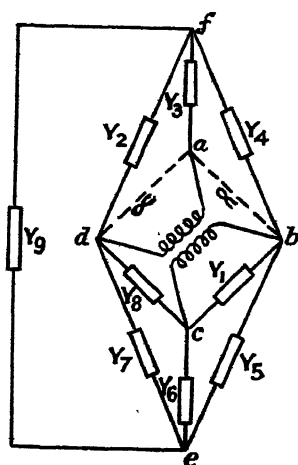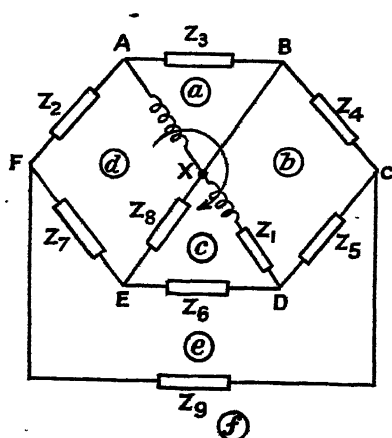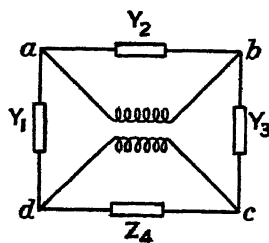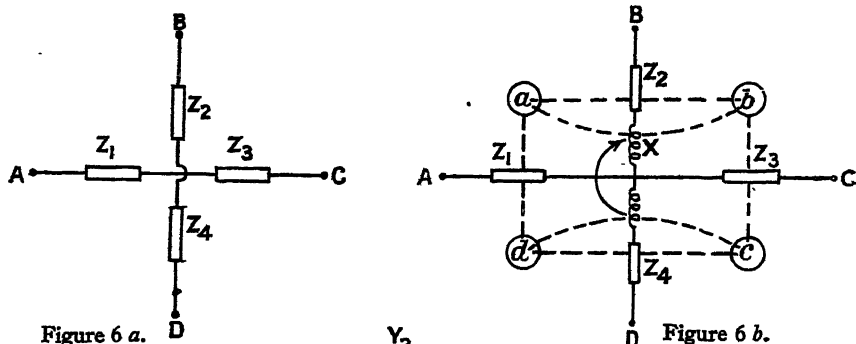
Figure 8 c.

Figure 8 b.

that this case can be dealt with by one ideal transformer with two windings only. Figures 8 b and 8 c show alternative solutions, obviously without exhausting the numerous possibilities.

The artifice of the fictitious junction point here introduced is of some further theoretical interest. As is well known, reciprocal diagrams of the type discussed occur also in graphical statics as reciprocal stress diagrams to pin-jointed frameworks. Frameworks with non-planar lay-out diagrams could not be solved by the original method. This difficulty was overcome by F. Schur (1897) by the introduction of fictitious pin-joints. Whilst, however, such a fictitious joint may be introduced into the lay-out of the framework without upsetting the stress distribution, the application of the analogous artifices in the case of the electrical network requires the special measures discussed above.

(e) *The method of the transferred terminal*

Certain cases of this method arise from an application of the method of the fictitious junction point just discussed. It will be noted that in the example of figure 7 the effect of introducing the fictitious junction point is that now Z_1 is connected directly * to junction point B instead of A. Such a " transfer of the terminal " will always take place, for instance, when the two leads crossing each other contain only one element each—instead of the two assumed in figure 6.

However, the transfer method is not restricted to the presence of such crossing branches. Quite generally if (figure 9 a) A and B represent two junction points of the network, and Z_t an element fed from point A, we can join A to B
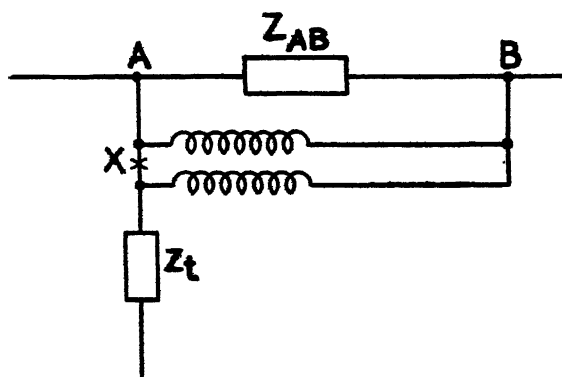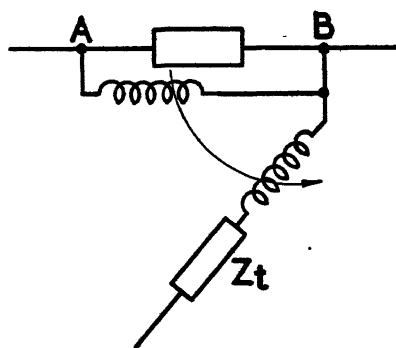
Figure 9 a.   Figure 9 b.

by the primary and secondary of an ideal transformer, as shown. We are then allowed to break the connection between primary and secondary at X—for on account of the transformer action no difference of potential can arise between the separated leads and, therefore, no change in the current through Z_t can take place. This current is, however, now supplied from B instead of from A; in addition, there flows in the primary of the transformer a current equal to that taken from B, in the direction from A to B. Hence, as far as the rest of the network is concerned, the current distribution remains unchanged. For the sake of clarity the change in the configuration of the network brought about by this step is illustrated again in figure 9 b.

* It is true that in the diagram shown, the transformer coil is interposed between the terminal of B and A ; however, the diagram could obviously have been drawn with the order of these two elements in branch 1 inverted.

This method, which can be described as a generalization of the dual to Julia's method, is illustrated in figure 10. Instead of eliminating a single branch from the original network we eliminate at one time several branches which together form one or more subsidiary networks.

Suppose we select in the original network a closed mesh containing the junction points (in this order) MNPQ .... and we join M to N, by the primary of an ideal two-winding transformer, similarly N to P, and so on. We are then allowed to join consecutively all the secondaries of these transformers so as to form a closed mesh. The mesh of the secondaries forms a kind of " parallel " mesh to the mesh formed by the primaries. Let us label the junction points of this second mesh M'N'P'Q' .... It is obvious that the potential difference between any one pair of these latter points is a replica of the potential difference of the corresponding unprimed points. For this reason any branch extending
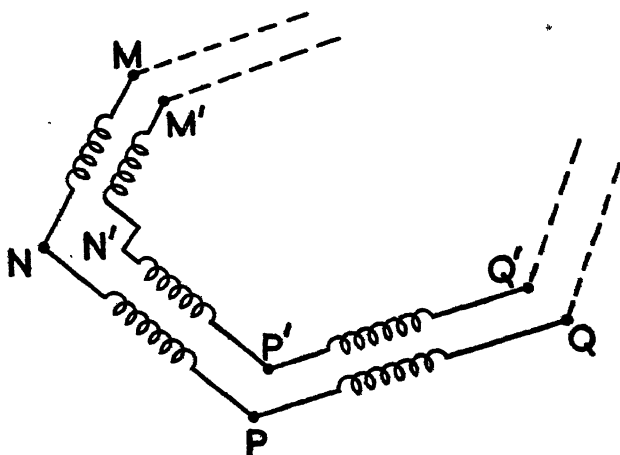
Figure 10.

between any pair of the unprimed junction points can be transferred to the corresponding primed points on the parallel mesh without altering the conditions of its working. The current drawn by such a branch still leaves the original network at the old junction points as currents which feed the primaries of the transformers concerned. Hence we have here a very simple means by which we can split the original network into two subsidiary networks which are only coupled by transformer action.

It is easily seen that the transformers need not actually form a closed mesh: one of the sides of the mesh may be left out. Furthermore, the mesh selected may—as was the case before (figure 5 *a*)—contain branches of infinite impedance.

This concludes the description of the various methods of network transformation which, it is felt, covers all needs arising in practice. No attempt has been made to cover the subject exhaustively.

Application

The possibility of constructing without restriction a network that is physically dual to a given network is in itself only of academic interest. It may, however,

become of immediate practical interest in connection with the increasing use of electro-mechanical analogies. As shown elsewhere (e.g. Bloch, 1945), we can study the action of an electro-mechanical system if we connect the electrical part of this system to a network which is the electrical analogue of the mechanical part of this system; this reduces the analysis to that of a single electrical network. However, this possibility exists only when we use the correct type of analogy: the direct one for the electrostatic type of coupling element and the inverse one for the electromagnetic type of coupling element. For the reasons already mentioned in the introduction to this paper, this method would break down— but for the methods here described—if we had to use the direct analogy in a case where the mechanical layout, and therefore the inverse analogue, were non-planar.

The methods of this paper enable us

(a) To draw a diagram of the direct analogue, and hence of the combined network, which we can use as a basis for further calculation or consideration. In this case it is obvious that we need not worry about the performance of those ideal transformers; they are ideal by definition.
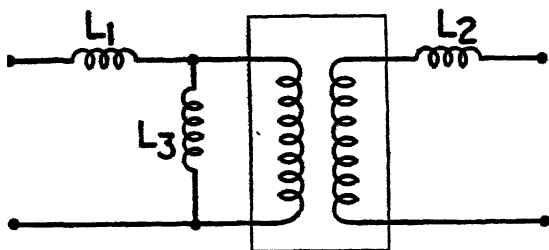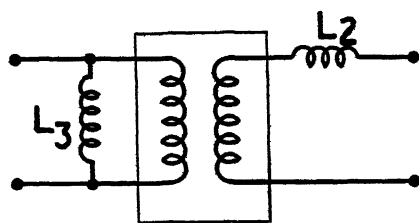
Figure 11 *a*.   Figure 11 *b*.

(b) To set up an actual circuit, say, on a network analyser which will enable us to study directly by measurement the performance of the electrical circuit or of the underlying mechanical system. In this case the ideal transformers have to be replaced by real ones. As real transformers do not behave exactly like ideal ones, errors may arise from this, and it is here that the possibility of choosing from a large number of different methods becomes of value; it may be possible then to select a method in which the imperfection of the real transformers are innocuous.

Figure 11 (a) shows a well-known equivalent circuit for a real transformer, i.e. a theoretical conception which explains the behaviour of a real transformer as arising from the combination of an ideal transformer with certain inductances: leakage inductances L_1 and L_2, and shunt inductances L_3. It is perhaps worth while to point out that the distribution of the leakage inductance on the primary and secondary side of the transformer is to some extent arbitrary, and depends on the assumption made as to the magnitude of the ratio of the ideal transformer. By a suitable assumption it is possible to have all the leakage inductance concentrated on the primary or on the secondary side. We arrive thus at an equivalent circuit such as shown in figure 11 *b*.

The error caused by the leakage inductance is of course small in all those

cases where the impedance of this additional inductance is small compared with all the other impedances in series with it. Similarly the error caused by the presence of the shunt inductance will be small when its admittance is small compared with the other admittances across which it is connected. In certain cases it will be possible to reduce these errors actually to zero—at least as far as the rest of the network is concerned. Suppose the transformer coil is connected in series with another inductance larger than the leakage inductance. Obviously we need then only reduce this other inductance by the amount of the leakage inductance in order to arrive at a combination which to the rest of the network
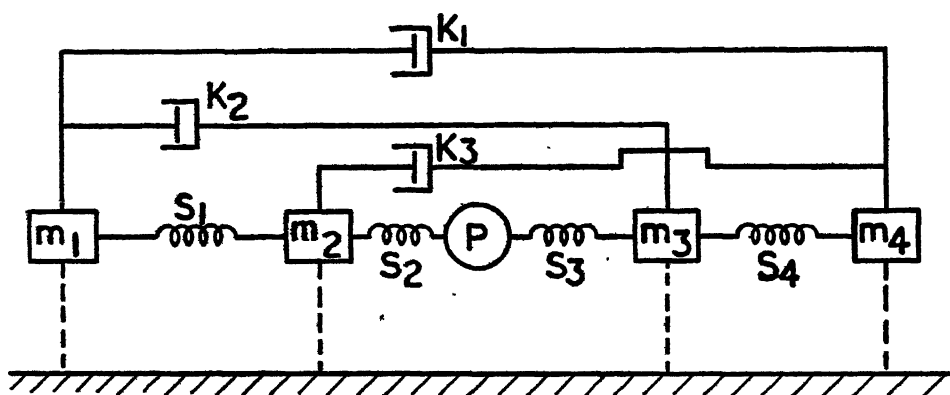
Figure 12.
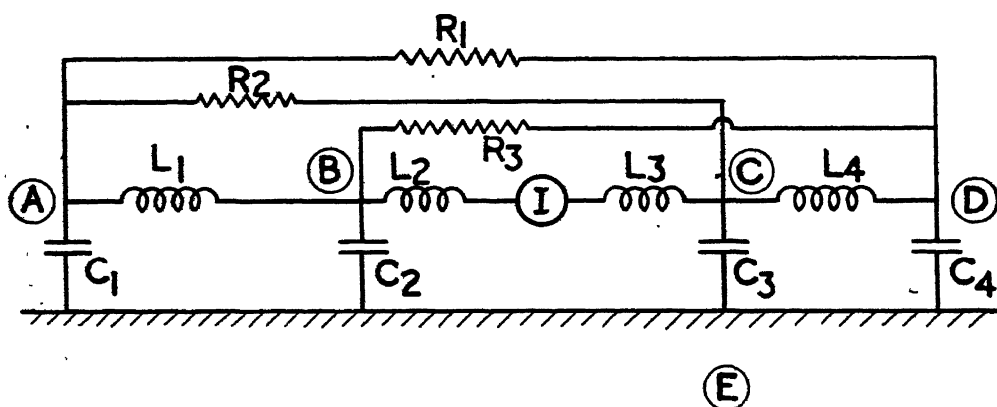
Figure 13.

looks exactly like the ideal combination. Similarly we can deal with the case of a shunt inductance in parallel to another inductance. The equivalent circuit of figure 11 b, in which all the leakage inductance is concentrated on one side, is specially useful for dealing with such cases. The combinations just referred to are wanted in the physical dual circuit; this means that when planning the original equivalent planar network we have to look ahead for their dual counterparts. Hence, for instance, when we wish to eliminate entirely the error caused by the leakage inductance we have to look for an equivalent network in which one transformer coil is placed in parallel to a capacity of sufficient magnitude.

Similarly, we have to place one transformer coil in series with a capacitative impedance of sufficient magnitude when we finally wish to eliminate the error caused by the shunt inductance.

An application of the aforesaid principles and of the principles of this paper in general is illustrated in figures 12 to 14. Figure 12 depicts a mechanical system consisting of a prime mover and of four masses which are inter-linked by an arrangement of springs and viscous couplings. The mode of presentation follows the convention which facilitates the drawing of the circuit of the electrical
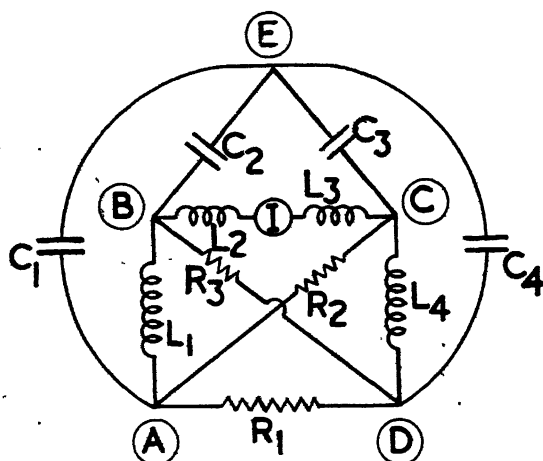
Figure 14.

inverse analogue. This circuit is shown in figure 13, and redrawn in a somewhat more familiar form in figure 14. From the latter figure it can be seen immediately that, for instance, a transformer could be placed in parallel to C_2, and thereby the terminal of C_1 transferred from E to B. Other " parallel " solutions are also obvious. A number of " series " solutions has already been given in figures $8a$ to $8c$; some of these (e.g., $8b$ or $8c$) enable us to place one transformer coil in series with a capacity—as stipulated above for the compensation of shunt inductance in the dual network.

REFERENCES

BARNES and GARDENER, 1942. *Transients in Linear Systems* (John Wiley), p. 49.
BLOCH, A., 1945. *J. Inst. Electr. Engrs.*, 92, Part I, 157.
CAUER, W., 1934. *Z. angewandte Math. Mech.*, 14, 349.
JULIA, R., 1939. *Bull. Soc. Franç. Electr.*, 9, 28.
RUSSELL, A., 1904. *Alternating Currents*, chapter XVII (Cambridge University Press).
SCHUR, F., 1897. *Math. Annal.*, 48, 142.
SIRE DE VILAR, H., 1901. *L'Eclairage Electrique*, 27, 252.
WHITNEY, HASSLER, 1931. *Proc. Nat. Acad. Sci.*, 17, 125.

ELECTRONIC STATES AND POTENTIAL CURVES
OF DIATOMIC MOLECULES

By J. G. VALATIN,
Budapest

ABSTRACT. A coordination rule of discrete molecular energy terms cannot be deduced from the non-crossing of the energy eigenvalue curves of the two-centre model. Clear distinction is to be made between the potential curves of unperturbed electronic states and the energy eigenvalue curves of the two-centre problem, the approximate existence of potential curves being the consequence of the averageability of the interactions with respect to the electronic motions which can also be influenced by the motion of the nuclei.

I N the coordination of molecular terms with atomic term combinations, it has for a long time been regarded as a quantum-mechanical principle that the molecular ground state dissociates into the atomic ground states, if these atomic states can form a molecular state of the corresponding symmetry (Herzberg, 1937; Hulthén, 1939). As has been pointed out several times by Schmid and Gerö, this coordination rule is inconsistent with strong spectroscopic arguments even in the case of some very important diatomic molecules, and the excited atomic states seem to play a much more important rôle from the point of view of chemistry. An adequate interpretation of all the spectral data leads to dissociation schemes for CO (Schmid and Gerö, 1937), CN (Schmid, Gerö and Zemplén, 1938), NO and N_2 (Schmid and Gerö, 1943), according to which the molecular ground state dissociates into higher atomic term combinations, and this applies also to a number of diatomic hydride and deuteride molecules (Gerö and Schmid, 1939, 1940).

The coordination rule mentioned resulted from a direct adaptation of the conclusions concerning the energy curves of the two-centre problem, whereas these need not have any consequences with respect to the discrete molecular energy levels. On the other hand, sharp distinction is to be made between the potential curves of unperturbed electronic states and the energy eigenvalue curves of the two-centre problem, the connection being much less close than has often been supposed. The motion of the electrons can also be influenced by the motion of the nuclei, and the wave-function factor characterizing an unperturbed electronic state is not identifiable with the eigenfunction of the two-centre problem in general.

We write the wave equation of the two-centre problem in the form

$$\{T_e + U\}\Phi_0(x_i, r) = E(r)\Phi_0(x_i, r), \qquad \dots\dots(1)$$

where x_i represents the coordinates of all the electrons, T_e symbolizes the kinetic energy of the electrons, U the interactions between all the particles; the fixed internuclear distance r occurs only as a parameter; and the energy eigenvalues

E and eigenfunctions Φ_0 of the problem result as functions of this parameter. As shown by Neumann and Wigner (1929), if certain symmetry conditions are fulfilled, two energy eigenvalues differing at some value of the continuously variable parameter r cannot coincide at another value of the internuclear distance, and energy curves of the same species do not cross. If in a certain approximation (for instance, in the approximation of the Heitler-London theory) two energy curves cross, then, taking into account the neglected terms, the energy values belonging to fixed values of r repel each other, and the curves avoid crossing. The figure shows how, with respect to the energy eigenvalue curves of the two-centre problem, this leads to the coordination rule mentioned. The corresponding conclusions have been applied also to the coordination of molecular and atomic terms (Hund, 1927).
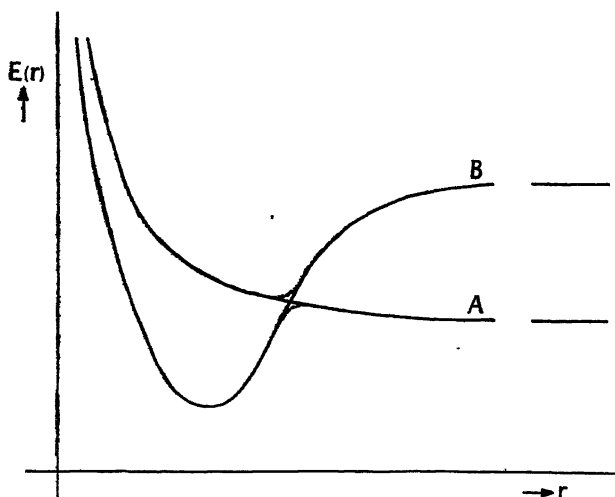
From the point of view of the molecule showing discrete energy levels, however, the nuclear distance cannot be regarded as a continuously variable parameter. The energy operator of the molecule, containing also a differential operator with respect to r, does not commute with the internuclear distance; energy and internuclear distance of the molecule are not even simultaneously measurable quantities in quantum mechanics. At a given value of the energy only the probability distribution of the r values is defined by the radial wave-function factor of the corresponding state. There is no sense in speaking of the dependence of energy on the nuclear distance, and the energy states of a molecule cannot even approximately be related to definite values of r. Accordingly, the corresponding conclusions cannot be transferred in this way from this model to the sequence of discrete molecular terms.

In the usual discussion of the diatomic molecule (Van Vleck, 1929; Kronig, 1930) it is supposed that the wave equation of the molecule can be separated into three parts, the neglected terms causing the perturbations observed in molecular spectra. One of these equations is the wave equation of the two-centre problem, the energy eigenvalue function of which enters into the equation describing nuclear vibrations, the potential curve of the molecule being defined in this way by the energy eigenvalues resulting at fixed nuclei. The potential function of

an unperturbed electronic state, however, can also be interpreted independently from these energy curves as the force law determining the nuclear motion which results from an averaging of all the interactions with respect to the electronic motions. This separation means a rather arbitrary division of the coupling between electronic and nuclear motions, which, in general, cannot be justified mathematically.

Using relative coordinates of the particles related to one of the nuclei, and using a reference system bound to the nuclei, the axes of which are given with respect to an inertia system by the Eulerian angles ϑ, ψ, ϕ, its z-axis showing the direction of the line joining the nuclei, we can write for the wave equation of a diatomic molecule,

$$\{T_r + T_e + U\}\Psi(x_i, r, \vartheta, \psi, \phi) = E\Psi(x_i, r, \vartheta, \psi, \phi), \qquad \ldots \ldots (2)$$

neglecting correction terms which can be ascribed to the finite mass connected with the origin of reference system. T_r now represents the relative kinetic energy of the nuclei, which involves the reduced mass, $\mu = M_1 M_2 / (M_1 + M_2)$, of the nuclei. We can restrict ourselves to singlet states, the inclusion of the spin in the usual way involving no complications.

Simultaneously with the energy equation, in the chosen reference system with respect to the total angular momentum of the molecule, we have

$$M^2\Psi = -\left\{\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial}{\partial\vartheta}\right) + \frac{1}{\sin^2\vartheta}\left(\frac{\partial}{\partial\psi} - iL_z\cos\vartheta\right)^2 - L_z^2\right\}\Psi = J(J+1)\Psi,$$
$$\ldots \ldots (3)$$

L_z denoting the operator corresponding to the component of the resultant electronic angular momentum in the direction of the nuclear axis. The rigorous validity of this equation, giving eigenvalues of a constant of motion, simultaneously with the wave equation, is to be especially emphasized, since this equation multiplied by the factor $\frac{h^2}{8\pi^2\mu}\frac{1}{r^2}$ often appears in molecular literature as the wave equation with only an approximate character of the rigid symmetrical top (Van Vleck, 1929; Kronig, 1930).

Denoting the corresponding components of the total and the resultant electronic angular momenta by M_x, M_y and by L_x, L_y, we can in first approximation neglect terms in the wave equation which represent a coupling between rotation and electronic motion, and which can be written in the chosen reference system as

$$-\frac{h^2}{4\pi^2\mu}\frac{1}{r^2}(M_x L_x + M_y L_y)\Psi; \qquad \ldots \ldots (4)$$

and we can take wave functions Ψ_Ω satisfying the restriction

$$-i\frac{\partial}{\partial\phi}\Psi_\Omega = L_z\Psi_\Omega = \Omega\Psi_\Omega. \qquad \ldots \ldots (5)$$

These can be factorized according to

$$\Psi_\Omega(x_i, r, \vartheta, \psi, \phi) = \Phi_\Omega(x_i, r)\,\Theta_\Omega^{J,M}(\vartheta, \psi, \phi). \qquad \ldots \ldots (6)$$

The function $\Theta_\Omega^{J,M}(\vartheta, \psi, \phi) = e^{iM\psi}e^{i\Omega\phi}Q_\Omega^{J,M}(\cos\vartheta)$ is a solution of equation (3),

and we get for the determination of $\phi_\Omega(x_i, r)$ the equation

$$\left\{\frac{h^2}{8\pi^2\mu}\left(-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{L_x^2+L_y^2-L_z^2}{r^2} + \frac{J(J+1)}{r^2}\right) + T_e + U\right\}\Phi_\Omega(x_i, r)$$

$$= E\Phi_\Omega(x_i, r). \qquad \ldots\ldots(7)$$

In accordance with the rigorous character of the rotational quantum number J, we have a natural way of separating the angular variables. The effect of the rotational coupling can be followed in the function of the quantum number J, the effect of the terms under expression (4) being negligible for $J \to 0$. We have another situation with respect to the coupling between nuclear vibration and electronic motion, which also is fully contained in equation (7). According to the purely dynamical character of nuclear vibration, there is no quantum number which retains its rigorous significance even in the case of strong coupling effects, and every approximate separation can by its nature have only a dynamical justification.

Supposing, with respect to the wave functions of an unperturbed electronic state,

$$\Phi_\Omega(x_i, r) = f(r)\,\Phi_e^{(\Omega)}(x_i, r) \qquad \ldots\ldots(8)$$

with

$$\int \overline{\Phi}_e^{(\Omega)}\Phi_e^{(\Omega)}dx_i \equiv 1, \qquad \ldots\ldots(8\,a)$$

we can multiply equation (7) by $\overline{\Phi}_e^{(\Omega)}(x_i, r)$ and integrate on the electronic coordinates. For the wave functions $f(r)$ characterizing nuclear vibration we get in this way the wave equation

$$\left\{\frac{h^2}{8\pi^2\mu}\left(-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{J(J+1)}{r^2}\right) + V(r) - E\right\}f(r) = 0, \qquad \ldots\ldots(9)$$

which yields the regular term system of an unperturbed electronic state of the molecule, with a potential function

$$V(r) = \int \overline{\Phi}_e^{(\Omega)}(T_e+U)\Phi_e^{(\Omega)}dx_i + \frac{h^2}{8\pi^2\mu}\int\overline{\Phi}_e^{(\Omega)}\left(-\frac{\partial^2}{\partial r^2} + \frac{L_x^2+L_y^2-L_z^2}{r^2}\right)\Phi_e^{(\Omega)}dx_i,$$

$$\ldots\ldots(10)$$

resulting from averaging the interactions.

The nuclear vibration is determined in this way by the motion of the electrons through the interaction of the particles; but, owing to the same interactions, we have to suppose that the motion of the nuclei can also influence the motion of the electrons with respect to which the averaging is performed. There is no mathematical reason to suppose that the functions $\Phi_e^{(\Omega)}$ can be identified with the eigenfunctions Φ_0 of the two-centre problem in general, and the first term in the form (10) of the potential function is identifiable with an energy eigenvalue function $E(r)$ of the two-centre problem. Wave equation (2) contains symmetrically the kinetic energies T_r of the nuclei and T_e' of the electrons, and it is not justifiable to neglect the motion of the nuclei simply in determining the electronic motion, this involving a generally arbitrary division of the coupling between nuclear vibration and electronic motion. The corresponding hypothesis would lead also to the experimental difficulties mentioned, and is not warranted by theory.

Potential curves can be defined only for the unperturbed electronic states. In passing from equation (7) to equation (9) we replaced the corresponding energy terms by their averages, neglecting coupling terms, to the effect of which can also be ascribed the homogeneous perturbations observable between the discrete molecular energy levels. The size of the displacements in the regular sequence of the levels is determined also in the case of homogeneous perturbations primarily by the actual position of the perturbing levels at a given value of J; while, in the interpretation of the effect, an important rôle should be ascribed also to the non-diagonal matrix elements of $T_e + U$—a point which seems to be left out of consideration in the interpretation of homogeneous perturbations by identifying the functions $\Phi_e^{(\Omega)}$ and Φ_0.

The solution of the molecular wave equation with the aid of the energy curves of the two-centre problem takes its origin from a paper of Slater (1927), where an analogous process is applied to the computation of the energy values of the helium atom, neglecting first the kinetic energy of one of the electrons and dealing with its coordinates as parameters. In the case of the much more complex problem of the molecule, this formal process may or may not give a correct approximation, and also the complications connected with an energy continuum are involved; for instance, the crossing curves of the figure as potential curves result in one discrete energy set, whereas the uncrossed energy curves of the two-centre problem would result in two energy sets.

In connection with the identification of the potential curves with the energy eigenvalue curves of the two-centre problem generally, a paper of Born and Oppenheimer (1927) is quoted, the citations often going farther, however, than the conclusions which can be drawn from this paper. In the second-order terms for the nuclear vibration a formal expansion process leads to the equation of a harmonic oscillator containing the curvature of the corresponding energy eigenvalue curve at the place of its minimum, but this does not involve any consequences with respect to an identification in the above sense.

Another formal solution with respect to the wave functions of the electronic states could be given by the composition and antisymmetrization of the wave functions of two atomic states. This also would necessarily contain the variable r and would still be multiplied by a function of r according to the normalization condition (8 a). Naturally, this formal approximation would not be suitable in general, but there may be cases when the potential curves resulting in this way give a still better approximation concerning the whole term system of the electronic state than the exact eigenvalue curves of the two-centre problem.

In the figure, if the crossing energy curves A and B represent the approximations of the Heitler-London theory, a perturbation calculation between the two energy curves leads in the two-centre problem to the non-crossing energy curves with wave functions

$$\Phi_0(x_i, r) = c_1(r)\,\Phi_0^{A}(x_i, r) + c_2(r)\,\Phi_0^{B}(x_i, r), \qquad \ldots\ldots(11)$$

where the factors $c_1(r)$, $c_2(r)$ may suffer a rapid variation between the values 0 and 1 near to the point of intersection. The approximations made by passing from equation (7) to equation (9) are also connected with the derivatives of $\Phi_e^{(\Omega)}(x_i, r)$ with respect to r; hence, in identifying these solutions $\Phi_0(x_i, r)$ of the two-centre problem

with the wave function of the electronic state, the terms neglected may become very significant, and we cannot even be sure whether the further approximations made between the resulting discrete energy levels will give converging results at all. But, in the case of a too rapid variation of the corresponding function with r, we by no means get a good approximation of the energy values which could characterize an unperturbed term system of the molecule.

In a paper dealing with atomic collision processes, London (1932) arrives at "nonadiabatic" transitions just from the rapid variation of the factors $c(r)$ in function (11), following the corresponding calculations also in detail. Naturally, we cannot identify ourselves with the stand-point of this paper (particularly from the point of view of the discrete energy levels), which starts from the wave functions corresponding to the energy states of the two-centre problem, and ascribes physical significance first of all to these. A discrete energy level of the molecule may be determined by a potential curve which differs essentially from the energy eigenvalue curve of the "adiabatic" problem, and may show predissociation effects (which are of importance in chemical reactions), the transition probability from the stable state to the continuum being connected with the matrix elements of $T_e + U$ as well as with those of T_r.

The possibility of a "potential theoretical treatment" of some molecular problem, the investigation of a one-body problem with a potential curve, is not restricted to the model of fixed nuclei, but arises from the averageability of the interactions with respect to the electronic motions in consequence of the smaller mass of the electrons. These electronic motions may also be influenced by the nuclear motion, and the potential curves cannot be identified in general with the eigenvalue curves of the two-centre problem. Especially, it is by no means possible to infer from the behaviour of the energy curves of this model a co-ordination rule of the molecular energy levels, or to consider the corresponding coordination rule as supported by theoretical arguments.

ACKNOWLEDGMENTS

REFERENCES

Born, M. and Oppenheimer, R., 1927. *Ann. Phys., Lpz.* (4), **84**, 457.
Gerö, L. and Schmid, R. F., 1939. *Z. Phys.*, **111**, 588 ; 1940. *Ibid.*, **115**, 57.
Herzberg, G., 1937. *Chem. Rev.*, **20**, 145.
Hulthén, E., 1939. *Z. Phys.*, **113**, 126.
Hund, F., 1927. *Z. Phys.*, **42**, 100.
Kronig, R. de L., 1930. *Band Spectra and Molecular Structure* (Cambridge University Press).
London, F., 1932. *Z. Phys.*, **74**, 143.
Neumann, J. v. and Wigner, E., 1929. *Phys. Z.*, **30**, 467.
Schmid, R. F. and Gerö, L., 1937. *Z. phys. Chem.* (B), **36**, 105.
Schmid, R. F., Gerö, L. and Zemplén, J., 1938. *Proc. Phys. Soc.*, **50**, 283.
Schmid, R. F. and Gerö, L., 1943. *Csillagászati Lapok*, **6**, 101, 102.
Slater, J. C., 1927. *Proc. Nat. Acad. Sci.*, **13**, 423.
Van Vleck, J. H., 1929. *Phys. Rev.*, **33**, 467.

PHOTOCHEMICAL DECOMPOSITION OF CO

By R. F. SCHMID and L. GERÖ,

Budapest *

MS. received from J. G. Valatin and communicated by W. Jevons 18 July 1946

THE photochemical decomposition of the CO molecule was investigated by Faltings, Groth and Harteck (1938). These authors irradiated CO gas at atmospheric pressure with a xenon lamp, the radiation consisting chiefly of the 1470 and 1295 A. resonance lines of xenon. In a path of 30 mm. length the CO gas absorbed about 70% of the incident light. Only the shorter wave-length was effective, causing CO_2 formation with a quantum efficiency of about unity; the 1470 A. line does not give rise to any CO_2 formation. Decomposition of CO was presumed to be the primary process; from this it was concluded that 219 kcal. (corresponding to the shorter wave-length) must be an upper limit and 193 kcal. (corresponding to the longer wave-length) probably a lower limit for the dissociation energy of the CO molecule.

As Gaydon and Penney (1942) have pointed out, no continuous absorption is known in the CO spectrum in the neighbourhood of the energy level of the 1295 A. xenon line; only the discrete bands of the Fourth Positive CO system lie in this part of the spectrum, and the $P(35)$ line of the (10,0) band coincides roughly with the xenon line. But, as the excitation of such a high rotational term of the ground state at room temperature is very improbable, they conclude that there is an absorption in the (13,2) band of the Fourth Positive system. If this is so, the primary process would then be:

$$CO(x^1\Sigma) + h\nu = CO(A^1\Pi),$$

$A^1\Pi$ being the upper state of the Fourth Positive band system. The mean lifetime of the CO molecule in the discrete $A^1\Pi$ terms cannot be much less than about 10^{-8} sec. In this way at least a hundred collisions take place before a quantum of fluorescence radiation could be emitted at room temperature and atmospheric pressure; these collisions with unexcited CO molecules would give rise to the formation of CO_2:

$$CO(A^1\Pi) + CO(x^1\Sigma) \rightarrow CO_2 + C.$$
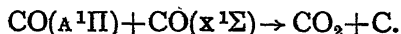
No conclusion in respect of the CO dissociation energy is available in this case.

Now the above assumption of Gaydon and Penney concerning the primary photochemical process is quite impossible. As the lines of the $A^1\Pi - x^1\Sigma$ band system of CO are rather sharp, the half-value width of the upper-state rotational terms cannot be greater than, say, 0.1 cm.$^{-1}$; with at least this accuracy the wave-number of a CO band line should then coincide with the wave-number of the absorbed xenon line. The extrapolated wave-numbers of the CO $A^1\Pi - x^1\Sigma$ (10,0) band lines in the neighbourhood of the xenon line are (Gerö, 1936; Schmid and Gerö, 1936; Gerö et al., 1927):

$P(35) = 77213.6$ cm.$^{-1}$,	$Q(37) = 77227.0$ cm.$^{-1}$,	$R(40) = 77207.1$ cm.$^{-1}$,
$P(36) = 77170.9$ cm.$^{-1}$,	$Q(38) = 77184.5$ cm.$^{-1}$,	$R(41) = 77164.1$ cm.$^{-1}$.
	$Q(39) = 77141.1$ cm.$^{-1}$,	

* It is much regretted that both authors have died since the preparation of the first draft of this paper.

The wave number of the xenon line is 77185·9 cm.$^{-1}$, calculated from the term values of Bacher and Goudsmit (1932); this wave number is nearest to the wave number of the $Q(38)$ line, but the accuracy of the coincidence is by no means sufficient.

The rotational quantum number belonging to the lower state of the $Q(38)$ line is $J = 38$. The probability of a state with such a high rotational quantum number is very low. Figure 1 contains the percentage probabilities of the rotational terms of the CO $x^1\Sigma(v=0)$ state at 300° K., evaluated according to the Boltzmann law:

$$100 \frac{(2J+1)e^{-\frac{hc}{kT}BJ(J+1)}}{\sum_J (2J+1)e^{-\frac{hc}{kT}BJ(J+1)}}.$$

The maximum probability lies at $J = 7$; at $J = 38$ its value is less than 0·0001%. The probability of the $x^1\Sigma(v=2)$ state is even less; besides, no lines of a (13,2) (extrapolated) band lie in the spectral region of the 1295 A. xenon line, the head of this red-degraded band being at 77059 cm.$^{-1}$.
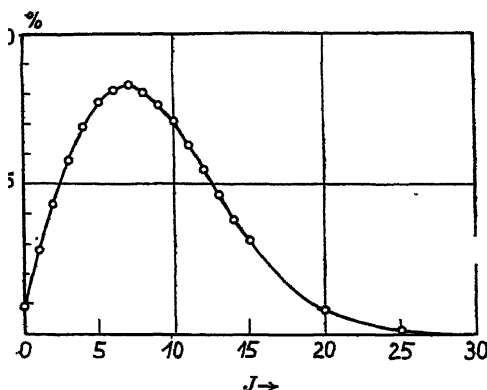
Figure 1.
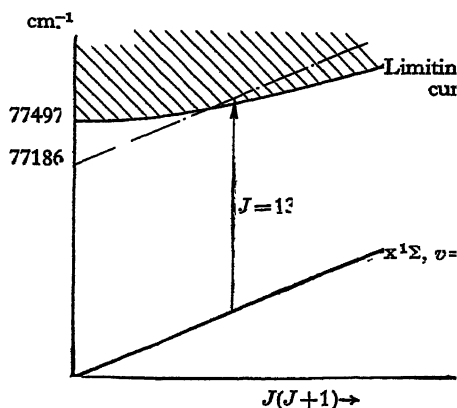
Figure 2.

But, furthermore, a third reason makes the absorption of the xenon line by a single band line quite impossible. Harteck and Oppenheimer (1942) have observed that the lines in the radiation of the xenon lamp are strongly broadened, on account of the great intensity of the current used. Now these broad lines cannot be absorbed up to 70% by discrete band lines, but only by a region of continuous absorption. In this way the primary process is an absorption whose upper state is a dissociation-continuum.

An explanation of the results of Faltings, Groth and Harteck without any contradiction is only possible (Schmid and Gerö, 1937) on the basis of the dissociation scheme:

$$CO(x^1\Sigma) \rightarrow C(^5S) + O(^3P) - 89620 \text{ cm.}^{-1}.$$
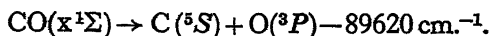
This dissociation scheme is founded on some predissociation effects, one of which has a limiting curve (Gerö, 1926) at 77497 ± 44 cm.$^{-1}$. Figure 2 contains this limiting curve. The predissociation which takes place along this curve is caused by molecular states belonging to the $C(^1S) + O(^3P)$ atomic term combination. The energy of the 1295 A. xenon line, 77185·9 cm.$^{-1}$, is not sufficient to raise a

CO molecule from its $x^1\Sigma(v=0, J=0)$ ground level over the limiting curve; but a considerable part of the molecule has at room temperature a rotational energy of some hundred cm.$^{-1}$. In figure 2 the rotational term sequence of the CO $x^1\Sigma$ ($v=0$) state is drawn against $J(J+1)$ as abscissa. At the height of 77185·9 cm.$^{-1}$, a parallel with this term sequence is drawn, which intersects the limiting curve in the neighbourhood of $J=13$. Hence a CO molecule in the $x^1\Sigma$ ($v=0$) ground state with $J \geqslant 13$ has quite enough rotational energy to dissociate on absorbing the 1295 A. xenon line. About 20% of all CO molecules are in states with $J \geqslant 13$ at room temperature; these molecules, absorbing the xenon radiation, rise in the dissociation continuum above the limiting curve, and as the mean life-time of this state is less than 10^{-13} sec., i.e., roughly one hundredth of the time interval between two collisions, the unit quantum efficiency of the dissociation process is quite well explicable.

From the $C(^1S) + O(^3P)$ atomic term combination, only triplet molecular terms arise. Transitions between the singlet ground state of the CO molecule and these triplet states are intercombination transitions. In the band spectrum of CO some intercombination band systems are known; one of them, the $a^3\Pi - x^1\Sigma$ (Cameron) band system, was photographed with fairly high intensity both in emission (Gerö, 1938) and absorption (Gerö, 1937). A number of perturbation and predissociation phenomena, caused by triplet states, were also observed in singlet states. For example, the predissociation in the $a^1\Pi$ molecular state at 77497 cm.$^{-1}$ is also caused by a triplet state; the intensity drop at this energy height is not very large and extends only over a few rotational levels of the $v=9, 8$ and 7 vibrational terms. Again, bands with $v'=10$ have rather high intensity. It follows that only above and close to the limiting curve is the probability of a transition without radiation sufficiently great to cause observable intensity diminutions. This is also the reason why no continuous absorption occurs at this energy level in the absorption spectrum of CO (see e.g. the absorption plates of Hopfield and Birge quoted in Bonhoeffer and Harteck, 1933). The effective region above the limiting curve is in fact too narrow, and the transition probability too small, to cause an observable absorption on plates taken with rather low resolving power.

Although the energy height of the ineffective 1470 A. xenon line—8·38 ev.— is about 1·5 ev. larger than the dissociation energy of CO, there is no atomic term-combination in the immediate neighbourhood of this energy height. Therefore, the CO molecule cannot absorb this wave-length. In this way no lower limit of a dissociation energy can be established by the non-appearance of a photochemical effect; a photodissociation possible on energy grounds only takes place if there is in the energy level of the absorbed radiation a continuum in which the transition probability from the ground level is large enough to give rise to considerable absorption.

REFERENCES

BACHER, R. F. and GOUDSMIT, S., 1932. *Atomic Energy States* (New York).
BONHOEFFER, K. F. and HARTECK, P., 1933. *Grundlagen der Photochemie*, p. 132.
FALTINGS, K., GROTH, W. and HARTECK, P., 1938. *Z. phys. Chem.* (B), 41, 15.
GAYDON, A. G. and PENNEY, W. G., 1942. *Nature, Lond.*, 150, 406.
GERÖ, L., 1936. *Z. Phys.*, 100, 374; 1938. *Ibid.*, 109, 204.
GERÖ, L., HERZBERG, G. and SCHMID, R. F., 1937. *Phys. Rev.*, 52, 467.
HARTECK, P. and OPPENHEIMER, F., 1932. *Z. phys. Chem.* (B), 16, 77.
SCHMID, R. and GERÖ, L., 1936. *Z. Phys.*, 101, 343; 1937. *Z. phys. Chem.* (B), 36, 105.

DISCUSSION

Dr. A. G. GAYDON. The comments by Schmid and Gerö on Gaydon and Penney's interpretation of the photodissociation of CO read plausibly, but are at variance with the direct experimental evidence on the absorption spectrum. This has been examined by Leifson (1926), using 15 mm. of gas at 1 atm. (i.e. half the path length used by Faltings, · Groth and Harteck). His description and published spectrogram show clearly that there is no general absorption of appreciable strength in this region ; since the (10,0) band is about the same strength as the (9,0) and (11,0) bands * it is most unlikely that there is a localized region of continuous absorption around 1295 A. Schmid and Gerö mention a *half-value* width of, say, 0·1 cm.$^{-1}$ for the band lines. This may be correct, although pressure effects might increase it somewhat at 1 atm., but it is the *overall* width, i.e. the width over which absorption is appreciable, with which we are concerned. For absorption in the main resonance system by 30 mm. of gas at 1 atm. this must be very much greater. Indeed since the lines of the bands are spaced several cm.$^{-1}$ apart it would not have been possible for Leifson, who used low resolving power, to record the strong absorption which he in fact observed if the overall line width had not been considerable. The experimental evidence clearly proves that the only strong absorption by CO between 1250 and 1400 A. is due to the Fourth Positive band system, and any interpretation of the photodissociation must start from this established fact.

It is now generally realized that the low value for the dissociation energy of CO favoured by Schmid and Gerö cannot be reconciled with equilibrium measurements of the heat of sublimation of carbon, which indicate a very much higher value than the 74 k.cal/mole required by Schmid and Gerö's theory. This has recently been discussed at some length (Gaydon, 1946).

REFERENCES

GAYDON, A. G., 1946. *Dissociation Energies and Spectra of Diatomic Molecules* (London : Chapman and Hall, in the press).
LEIFSON, S. W., 1926. *Astrophys. J.*, 63, 73.

* In Leifson's spectrogram there is some N_2 absorption superposed on the (8,0) and (9,0) bands of CO.

NOTE ON THE CALCULATION OF OPTICAL CONSTANTS

BY D. J. PRICE,

Walthamstow

THE calculation of the optical constants of an absorbing medium from the polarization produced in a beam of light reflected from the surface presents certain difficulties due to· the somewhat cumbersome equations usually employed.

A new equation more suitable for routine computing is here developed and discussed with reference to previous work.

Refraction in an absorbing medium may be characterized by a complex refractive index, $N = n + ik$. This results in a complex angle of refraction χ, corresponding to a real angle of refraction ϕ, for by Snell's law

$$\sin\phi = N \sin\chi. ·$$

Further, if the incident light consists of two equal components in phase, and polarized in directions parallel and perpendicular to the plane of incidence, then

after reflection the phase difference Δ and the ratio of the two amplitudes ρ will be given by Fresnel's equations as

$$-\rho e^{i\Delta} = \frac{\cos(\phi+\chi)}{\cos(\phi-\chi)}.$$

It is usual to measure Δ by restoring the phase difference by means of some type of compensator and then determining the azimuth of the resulting plane polarized light with respect to the plane of incidence. If this angle of azimuth be designated ψ, then

$$\rho = \tan\psi.$$

Combining the three equations, we have, after a simple reduction,

$$N^2 = \left(\frac{1-Z}{1+Z}\right)^2 \sin^2\phi\tan^2\phi + \sin^2\phi,$$

where

$$Z = e^{i\Delta}\tan\psi$$
$$= (\cos\Delta + i\sin\Delta)\tan\psi.$$

It is from this that most workers have proceeded. Drude (1889 to 1898), to whom the first analysis is due, developed two alternative methods.

First, for an approximate treatment, the $\sin^2\phi$ term may be neglected. This enables the root to be taken on both sides and therefore yields a simple expression for N. There are two objections to this approximation for modern work: (a) it is more inaccurate than is allowable with more recent experimental techniques, and (b) the theory of dispersion of metals involves N^2 and not N itself.

Secondly, Drude introduces three new, intermediate quantities defined by the equations :—

$$\tan Q = \sin\Delta\tan 2\psi,$$
$$\cos P = \cos\Delta\sin 2\psi,$$
$$S = \sin\phi\tan\phi\tan\tfrac{1}{2}P.$$

Taking $N^2 = A + iB$, he obtains, after some reduction,

$$A = S^2\cos 2Q + \sin^2\phi,$$
$$B = S^2\sin 2Q.$$

This two-stage formula, although leading to a fairly simple form, entails quite a lengthy computation.

A simple and *direct* formula may be developed as follows :—

$$N^2 = \left(\frac{1-Z}{1+Z}\right)^2 \sin^2\phi\tan^2\phi + \sin^2\phi$$

$$= \frac{2\sin^2\phi\tan^2\phi}{\tfrac{1}{2}\left(Z+\dfrac{1}{Z}\right)+1} + \tan^2\phi.$$

But

$$\tfrac{1}{2}\left(Z + \frac{1}{Z}\right) + 1 = \frac{\cos\Delta}{\sin 2\psi} + \frac{\sin\Delta}{\tan 2\psi}.$$

Hence, if we put

$$X = \frac{\cos\Delta}{\sin 2\psi} + 1$$

and

$$Y = \frac{\sin\Delta}{\tan 2\psi},$$

we derive
$$A = \tan^2\phi - 2\sin^2\phi\tan^2\phi\left(\frac{X}{X^2+Y^2}\right),$$
$$B = 2\sin^2\phi\tan^2\phi\left(\frac{Y}{X^2+Y^2}\right).$$

Further,
$$X^2 + Y^2 = (\operatorname{cosec}2\psi + \cos\Delta)^2.$$

Therefore,
$$A = \tan^2\phi - \frac{2\sin^2\phi\tan^2\phi\,(\cos\Delta + \sin 2\psi)}{\sin 2\psi(\cos\Delta + \operatorname{cosec}2\psi)^2},$$
$$B = \frac{2\sin^2\phi\tan^2\phi\sin\Delta}{\tan 2\psi(\cos\Delta + \operatorname{cosec}2\psi)^2}.$$

This form is well suited to a series of computations based on observations made at different wave-lengths with a constant angle ϕ. In such a case the calculations may be set out in the following standard form :—

$$\phi = 73^\circ 0', \qquad\qquad D = \tan^2\phi = 10\cdot698,$$
$$\Delta = 68^\circ 2\tfrac{1}{2}', \qquad\qquad C = \log(2\sin^2\phi\tan^2\phi)$$
$$2\psi = 59^\circ 38', \qquad\qquad\quad = 1\cdot29155.$$

		logs	*differences*
	C	$= 1\cdot29155$	$0\cdot92051 \quad \ldots\ldots(a)$
		$0\cdot37104$	
$\operatorname{cosec}2\psi \ =$	$1\cdot15900$	$\uparrow \times 2$	
sum	$1\cdot53292 \longrightarrow$	$0\cdot18552$	
$\cos\Delta \ =$	$0\cdot37392$		
sum	$1\cdot23673 \longrightarrow$	$0\cdot09228$	$0\cdot15636 \quad \ldots\ldots(b)$
$\sin 2\psi \ =$	$0\cdot86281 \longrightarrow$	$\bar{1}\cdot93592$	
	$\log\sin\Delta$	$= \bar{1}\cdot96729$	$\bar{1}\cdot73512 \quad \ldots\ldots(c)$
	$\log\tan 2\psi$	$= 0\cdot23217$	

		antilogs			
$a+c$	$0\cdot65563 \longrightarrow$	$4\cdot525$	$= B$	$= 2nk$	
$a+b$	$1\cdot07687 \longrightarrow$	$11\cdot936$			
		$10\cdot698$	$= D$		
diff.		$1\cdot238$	$= -A$	$= k^2 - n^2$	

It was found in practice that the substitution of this method of calculation for the previously noted formula of Drude involved a reduction of computing time by a factor of about two.

REFERENCES

Drude, 1889. *Ann. Phys., Lpz.*, 36, 532 ; 1890. *Ibid.*, 39, 481 ; 1898. *Ibid.*, 64, 159.

ULTRA-VIOLET ABSORPTION BAND-SYSTEMS OF SnSe AND SnTe

By E. E. VAGO and R. F. BARROW,

Physical Chemical Laboratory, University of Oxford

ABSTRACT. The study of the absorption spectra of SnSe and of SnTe (Walker, Straley and Smith, *Phys. Rev.* **53**, 140 (1938) ; Barrow and Vago, *Proc. Phys. Soc.* **56**, 78 (1944)) has been extended to cover the ultra-violet region down to 2100 A. New systems of both molecules have been analysed vibrationally. The results, including those derived from the D–X systems already known, are summarized in the following table :

Molecule	State	ν_e	ω_e	$x_e\omega_e$	$y_e\omega_e$	D (ev.)
SnSe	F	(47850)	(290)	—	—	—
	E	30738·9	196·6	−0·77	−0·0016	0·7
	D	27549·6	225·1	−0·69	—	1·1
	X	0	331·2	−0·73ₑ	—	4·5
SnTe	G	(47260)	(230)	—	—	—
	F	44033·5	229·7	−1·25	−0·003	—
	E	(28000)	(150)	—	—	0·6
	D	25444·3	179·1	−0·40	—	0·9
	X	0	259·5	−0·50	—	4·1

§1. INTRODUCTION

THE results of some work on the emission and absorption spectra of SnSe and SnTe were reported in a discussion on band-spectra (Barrow, 1944). However, the low-wave-length limit of these absorption studies was about 3700 A., and there were several reasons which led us to suspect the existence of at least one further absorption system (E–X) of both molecules at shorter wave-lengths. For, not only were there well-known E–X systems of SnO and SnS, but also bands which apparently belonged to these systems had already been identified in the emission spectra of SnSe and SnTe (Barrow, 1940; Barrow and Vago, 1943).

We decided therefore to examine the absorption of SnSe and SnTe vapours in the region 2100–3800 A. (at temperatures of the order of 1000° c.). This proved to be more fruitful than we had expected.* We have been able not only to confirm the existence of the E–X systems of these molecules (Barrow and Vago, 1944), but also, for SnSe, to carry out the vibrational analysis of the rather extensive band-system with v'-assignments going from 0 to 29: this leads to a fairly reliable estimate of the dissociation energy for the E state of SnSe. Unfortunately the absorption spectrum of SnTe in the corresponding region is rather complicated, and we have failed to produce a convincing vibrational scheme: for the E–X system of this molecule we are certain, therefore, only of the approximate position of the E level relative to the ground state.

---

* While this work was being prepared for publication a note on the same subject was published by D. Sharma (*Nature*, **157**, 663 (1946)). His results for SnSe and SnTe appear to be similar to ours, but his interpretation differs in detail. Further comment may be postponed until a full account of his work is available.

At shorter wave-lengths than the E–X systems we have found two new systems of SnTe and one of SnSe. . Of these, the F←X system of SnTe was quite well developed, and the vibrational analysis presented few difficulties. We have obtained only fragmentary information about the G←X system of SnTe and the F←X system of SnSe, as the low-λ portions of these systems lie beyond the limit of the present measurements. Corresponding systems of SnO and SnS may be expected in the vacuum ultra-violet region (1600–2000 A.), but no molecule of the group—apart from the first member, CO—has yet been studied in this region: there are several indications that such work would be worth while.

The results obtained with SnSe and SnTe suggested an examination of the absorption spectra of the analogous lead compounds, PbO, PbS, PbSe and PbTe, in the same region. This we have done, and have found new systems of the last three compounds: the results will be reported in a later paper. The experimental methods have been described in earlier papers in this series; we therefore proceed to consider the results.

§ 2. RESULTS FOR SnSe

Ultra-violet absorption bands attributed to SnSe have been found in two regions, 4150–2730 A. and 2200–<2070 A. These bands have been assigned to three systems which are discussed below. Plates taken at somewhat lower pressures of SnSe than are required for the best development of the bands in the region 4150–2730 A. show what appears to be continuous absorption at λ<2730 A.

(i) The D–X system: $\lambda\lambda$ 3410–4150 A.

A vibrational analysis of this sytem, based on emission spectrograms, was given some time ago (Barrow and Vago, 1943). We have now confirmed this analysis by measurements of absorption plates taken in a first order of the 2·4-m. concave-grating instrument (linear dispersion 7·4 A./mm.). The positions of a number of new bands assigned to this system have been located. Both sets of measurements are summarized in a Deslandres scheme (table 1): improved values of the band-head constants derived from table 1 are given in table 7.

(ii) The E–X system: $\lambda\lambda$ 2730–3410 A.

A spectrogram of the bands assigned to this system, taken with a large quartz Littrow spectrograph (dispersion ~5 A./mm. at 3250 A.), is reproduced in figure 1. Significant are (1) the long, rather regular, progression of bands in the region 2850–3050 A., (2) the sharpness of the band-heads around 3250 A., and (3) the overlap with bands of the D–X system at about 3400 A. Of these features, (1) gives immediately the order of magnitude of ω_e', since these bands are most likely to constitute a progression with $v''=0$, and (2) provides an approximate value for ν_e, since the diffuseness of the outlying band-heads may be attributed to the vibrational isotope effect (Barrow and Vago, 1943). Since, in addition, the ground-state vibrational terms for SnSe are already known quite accurately, it will be clear that there cannot be many different ways of assigning plausible v', v'' values to the bands. We have found only one such way (table 2). This analysis leads to the values of the constants listed in table 7. It should be

Figure 1.   Bands of the D-X and E-X systems of SnSe in absorption (quartz Littrow spectrograph

(a) Sn Se; F←X

(b) Sn Te;
F←X
G←X

(c) Sn Te;
D⇌X
E⇌X

Table 1. Band-head data for the D⇌X system of SnSe

Wave-numbers, cm⁻¹; bold-face denotes bands measured both in absorption and in emission; large Roman, bands measured in absorption only; large Roman*, bands measured in emission only.

(In each cell: band-head wave-number over *wavelength in Å*. Asterisk = emission only.)

v'' →	0	1	2	3	4	5	6	7	8	9	10	11	12
	29249.4* / *3417.90*	28918.6* / *3457.00*											
	29041.5* / *3442.37*	28712.7* / *3481.78*											
	28820.0* / *3468.82*	28494.4* / *3508.46*				27185.6 / *3677.87*							
	28601.0* / *3496.38*	28270.2* / *3536.29*											
	28383.4* / *3522.18*	28054.4* / *3563.49*	27726.1 / *3605.69*	27399.2 / *3648.70*	27073.5 / *3692.60*								
	28165.0* / *3549.60*	27836.6* / *3591.37*	27507.2 / *3634.38*	27178.8 / *3678.30*	26853.5 / *3722.86*	26528.8 / *3768.42*		26106.1 / *3829.44*					
	27944.6* / *3577.49*	27615.2* / *3620.17*	27285.2 / *3663.95*	26957.7 / *3708.47*	26631.6 / *3763.88*	26309.5 / *3799.83*	25988.1 / *3846.83*				24709.2 / *4046.94*	24397.7 / *4097.60*	24083.2 / *4161.11*
	—	27391.6 / *3649.72*	27062.7 / *3694.18*	26736.1 / *3739.20*	26410.9 / *3785.24*	26086.8 / *3832.28*		25444.1 / *3929.07*	25124.9 / *3979.00*	24806.2 / *4030.12*	24490.4 / *4082.08*	24174.3 / *4135.46*	
	27166.6* / *3679.95*	27166.6 / *3679.95*	26839.0 / *3724.87*	26511.5 / *3770.88*	26185.7 / *3817.80*	25861.8 / *3865.62*	25540.1 / *3914.31*	25221.0 / *3963.83*	24902.0 / *4014.61*	24583.7 / *4066.59*			

Vibrational combination intervals (cm⁻¹) read between band heads:

Between adjacent v'' columns (horizontal, ≈ ΔG''): 330.0, 328.8, 325.6, 330.8, 329.0, 328.4, 329.4, 327.6, 328.3, 329.4, 330.0, 328.9, 326.9, 328.4, 327.5, 326.6, 327.5, 325.7, 325.3, 326.1, 325.2, 325.8, 324.7, 322.1, 324.1, 323.9, 321.4, 321.7, 319.2, 319.1, 319.0, 318.7, 318.3, 315.8, 311.5, 316.1, 314.5, 218.8, 223.4

Between adjacent v' rows (vertical, ≈ ΔG'): 207.9, 221.5, 219.0, 217.6, 218.4, 220.4, 223.6, 225.0, 205.9, 218.3, 224.2, 215.8, 217.8, 221.4, 220.4, 222.0, 222.5, 223.7, 220.0, 221.1, 222.0, 223.7, 219.3, 221.9, 222.7, 220.7, 219.1, 222.9, 222.5, 222.7, 225.0, 225.2, 224.6, 223.2

Table 2. Band-head data for the E–X system of SnSe

e indicates band also observed in emission.

* band possibly present, but overlaid by the 7,0 band of the D–X system.

The table is a Deslandres (band-head) array; the upper-right blocks labelled "v″ → 0" are continuations of the v″ = 0 column for higher v′. Each entry gives the band head in cm⁻¹ (roman) above the wavelength in Å (*italic*). "e" = also in emission.

v′	v″ = 0	v″ = 1	v″ = 2	v″ = 3	v″ = 4	v″ = 5	v″ = 6
29	35175 / *2842·1*						
28	35118 / *2846·7*						
27	35044 / *2852·7*						
26	34956 / *2859·9*						
25	34866 / *2867·3*						
24	34774 / *2874·9*						
23	34671 / *2883·4*						
22	34545 / *2893·9*						
21	34405 / *2905·7*						
20	34256·4 / *2918·31*						
19	34101·1 / *2931·60*						
18	33940·2 / *2945·50*						
17	33775·7 / *2959·84*						
16	33605·4 / *2974·85*						
15	33438·0 / *2989·74*						
14	33266·5 / *3005·15*	32930 / *3035·9*					
13	33080·7 / *3022·03*	32749·1 / *3052·61*					
12	32906·3 / *3038·05*	32576·0 / *3068·86*					
11	32728·5 / *3054·56*	32398·4 / *3085·68*					
10	32550·5 / *3071·26*	32221·4 e / *3102·63*					
9	32369·3 / *3088·45*	32042·0 e / *3119·99*		31387·5 / *3185·06*		30745·3 / *3251·60*	
8	32188·4 / *3105·81*	31860·1 / *3137·82*	31531·6 / *3170·51*	31206·5 e / *3203·54*		30562·4 / *3271·06*	30240·3 / *3305·90*
7		31676·2 / *3156·03*	31348·8 e / *3189·00*	31022·8 / *3222·51*	30697·0 / *3256·71*		
6		31488·8 / *3174·83*	31160·5 / *3208·26*	30834·8 e / *3242·15*	30510·6 / *3276·61*		
5		31298·8 / *3194·51*	30971·0 e / *3227·90*	30645·4 / *3262·18*	30320·1 e / *3297·19*		
4	31441·1 / *3179·64*	31109·7 e / *3213·51*	30782·1 e / *3247·69*	30456·2 e / *3282·42*	30128·6 e / *3318·16*	29806·4 / *3354·02*	
3	31251·6 / *3198·89*	30921·8 / *3233·03*	30591·5 e / *3267·95*	30264·8 / *3303·22*	29940·7 e / *3338·98*	29617·6 e / *3375·40*	
2		30732·3 / *3252·97*	30400·6 / *3288·46*	30073·8 / *3324·00*	29747·9 e / *3360·62*	29425·7 e / *3397·42*	29104·9 e / *3434·37*
1		30537·4 / *3273·72*	30208·1 / *3309·42*	29881·8 / *3345·55*	29557·7 / *3382·25*	29229·4 / *3420·25*	
0				29684·2 / *3367·83*	29360·0 / *3405·44*	*	

Horizontal (Δv″) differences between adjacent band heads in each v′ row

v′	0→1	1→2	2→3	3→4	4→5	5→6
14	327					
13	331·6					
12	330·3					
11	330·1					
10	329·1					
9	327·3					
8	328·3	328·5	325·1		322·1	
7		327·4	326·0	325·8		
6		328·3	325·7	324·2		
5		327·8	325·6	325·3		
4	331·4	327·6	325·9	327·6	322·2	
3	329·8	330·3	326·7	324·1	323·1	
2		331·7	326·8	325·9	322·2	320·8
1		329·3	326·3	324·1	328·3	
0				324·2		

Vertical (Δv′) differences down each column

v″ = 0 column: 57, 74, 88, 90, 92, 103, 126, 140, 149, 155·3, 160·9, 164·5, 170·3, 167·4, 171·5, 175·8, 174·4, 177·8, 178·0, 181·2, 180·9, … , 189·5

v″ = 1 column: 181, 173·1, 177·6, 177·0, 179·4, 181·9, 183·9, 187·4, 190·0, 189·1, 187·9, 189·5, 194·9

v″ = 2 column: 182·8, 188·3, 189·5, 188·9, 190·6, 190·9, 192·5

v″ = 3 column: 183·7, 188·0, 189·4, 189·2, 191·4, 191·0, 192·0, 197·6

v″ = 4 column: 186·4, 190·5, 191·5, 187·9, 192·8, 190·2, 197·7

v″ = 5 column: 188·8, 191·9, 196·3, 182·9

noted, however, that the expression for $G_v{'}$ derivable from table 7 is valid only up to $v' = 17$: for the representation of higher-v' levels, additional terms in higher powers than $(v+\frac{1}{2})^3$ would be necessary. The biggest uncertainty in the analysis attaches to the reality of the bands with $v' = 0$. The difficulties here are exactly analogous to those encountered in the analysis of the E–X system of SiS, in which there is a rather similar intensity distribution (Vago and Barrow, 1946). For .SnSe, however, there is the additional confirmation from the vibrational isotope effect that the scheme given in table 2 is correct.

(iii) The F←X system: $\lambda\lambda < 2070$–2200.

Only a part of this system has been observed (figure 2, table 3). Most of the bands appear to form a progression, probably with $v' = 0$. That this is a reasonable conclusion may be seen by comparing the intensity distribution with that for the F–X system of SnTe (figure 2). The wave-number differences then identify the absorbing molecule as SnSe, and lead to a provisional assignment of v'' values which can hardly be in error by more than ± 2 units and may well be correct. The estimate of the upper state frequency given in table 7 rests on measurements of only two weak bands: it accords satisfactorily with what might be expected by analogy with the F-state data for SnTe, but the figure given cannot be regarded as much more than a reasonable guess.

Table 3. Band-head data for the F←X system of SnSe

Wave-length (I.A.)	ν, cm.$^{-1}$	$\Delta\nu$	Provisional v', v'' values
2077·5	48119		1,0
		333	
2092·0	47786		1,1
		288	
2104·7	47498		0,1
		328	
2119·3	47170		0,2
		324	
2134·0	46846		0,3
		325	
2148·9	46521		0,4
		323	
2163·9	46198		0,5
		326	
2179·3	45872		0,6
		324	
2194·8	45548		0,7
2080·5	48050		?
2096·6	47681		?

§3. RESULTS FOR SnTe

The banded ultra-violet absorption spectrum of SnTe can also be divided into two regions: 3075–4200 A. and <2120–2360 A. There is some evidence of a region of continuous absorption at $\lambda < 3075$ A.

(i) The D–X system: $\lambda\lambda$ 3700–4200 A.

The vibrational structure of this system has been adequately discussed in earlier papers. Since, however, there is a region where the long-λ end of the E–X system overlaps the D–X system, we did make a particular attempt to sort

out the bands which might be assigned to high v' values in the D–X system (3680–3750 A.: figure 2, table 4).   We were not able to draw any definite conclusions.

Table 4.   Band-heads of the E⇌X system of SnTe

Wave-length (I.A.)	Wave-number, cm.⁻¹ (in vac.)	Intensity	Wave-length (I.A.)	Wave-number, cm.⁻¹ (in vac.)	Intensity	
3327·2	30046	–	3540·82	28234·0 *	1	
35·5	29972	–	44·51	204·6 *	2	
42·1	913	–	54·54	125·0 *	2	
50·4	840 *	–	59·13	088·8 *	1	
65·5	705 *	–	65·63	037·6 *	3	
74·1	629 *	–	76·66	27951·1 *	4 e	
80·3	575 *	–	78·35	937·5 *	1	
88·5	503	–	87·17	869·2 *	5 e	
95·6	442 *	–	91·88	832·7 *	2 e	
3403·8	371 *	–	98·34	782·7 *	7 e	
11·7	303	–	3611·31	682·9 *	6 e	
19·7	234 *	–	20·28	614·3 *	3	
27·5	167	–	26·04	570·5 *	2 e	
34·5	108 *	–	31·73	527·3 *	5	
43·9	029 *	–	42·15	448·5	2 e	
50·5	28973 *	–	45·07	426·5 *	4 e	
58·8	903 *	1	59·82	316·0 *	9 e	
68·0	827	1	63·49	288·6	1	
74·5	773 *	0	73·14	216·9	2	
83·1	702 *	2	76·36	193·1 *	2 e	
91·5	633 *	0	84·64	132·0 *	1	? D⇌X, 10,0
99·46	567·7 *	2	94·04	063·0 *	10 e	
3509·08	489·4 *	3	3706·53	26971·8 *	3	? D⇌X, 9,0
12·1	465 *	0	11·17	938·0 *	3 e	
14·6	444 *	0	18·38	885·8 *	4	? D⇌X, 10,1
16·9	426	0	29·03	809·0 *	10 e	D⇌X, 8,0
22·94	377·3 *	1	41·74	718·0 *	3	? D⇌X, 9,1
27·01	344·6 *	0	46·65	682·9 *	3 e	
31·05	312·1 *	1				
34·79	282·2	2				

e indicates a band also observed in emission.
* indicates a band which may be fitted into a v' progression.

(ii) The E–X system: $\lambda\lambda$ 3075–3720 A.

When bands of this system were first observed—in emission—an attempt was made to carry out a vibrational analysis (Barrow, 1940).   That only a tentative result could be given was attributed to the incompleteness of development of the system under these conditions.   The new absorption plates (figure 2) are much superior to the emission ones, and we are satisfied that the band-head measurements given in table 4 are reliable.   Nevertheless, we have not been able to assign the bands to any really plausible vibrational scheme.   One difficulty—that the long-λ end of this system overlaps the short-λ end of the D–X system—has been mentioned.   Others are that the value of $x_e''\omega_e''$ for SnTe (0·5 cm.⁻¹) is so small as to preclude the unambiguous assignment of v'' values from wave-number

differences alone, and that neither sequences nor progressions are particularly well marked. Over a considerable range—some 1000 to 2000 cm.$^{-1}$ to the short-λ side of the origin—a frequency difference of about 135 is common, suggesting that ω_e' may be about 150 cm.$^{-1}$. We have, however, only succeeded in assigning the bands to a scheme involving this frequency by assuming considerable and rather numerous upper-state perturbations, about which we have no independent information.

The vibrational isotope effect enables the system-origin to be fixed at 28000 ± 250 cm.$^{-1}$ (3570 A.), and although bands can be followed down to about 3075 A., we have no evidence that there is more than one upper electronic level of SnTe in the energy range corresponding to these wave-lengths.

(iii) The F←X system: $\lambda\lambda$ 2190–2360.

This system is bounded by two strong progressions (figure 2), and there can be little doubt about the correctness of the analysis given in table 5.

Table 5. Band-head data for the F←X system of SnTe

v'	$v''=0$	1	2	3	4	5	6
8	45759 / *2184·7*	[259] 45500 / *2197·1*					
	205	204					
7	45554 / *2194·5*	[258] 45296 / *2207·0*		44779·3 / *2232·48*			
	210			208·6			
6	45343·9 / *2204·68*			44570·7 / *2242·93*			
	213·9						
5	45130·0 / *2215·13*		44611·4 / *2240·88*		44100·0 / *2266·87*		
	217·6		217·4				
4	44912·4 / *2225·86*		44393·0 / *2251·91*	[255·5] 44137·4 / *2264·95*			
	221·0			216·4			
3	44691·4 / *2236·87*			43921·0 / *2276·11*			
	219·3						
2	44472·1 / *2247·90*	[259·6] 44215·5 / *2261·10*					
	224·9	224·1					
1	44247·2 / *2259·33*	[258·8] 43988·4 / *2272·62*	[259·1] 43729·3 / *2286·09*				
		229·4	227·9				
0	—	43759·0 / *2284·54*	[257·6] 43501·4 / *2298·07*	[256·1] 43245·3 / *2311·68*	[255] 42990 / *2325·4*	[258] 42732 / *2339·4*	[253] 42479 / *2353·4*

$v'' \rightarrow$ 0 1 2 3 4 5 6

(iv) The G←X system: $\lambda\lambda<2120$–2190.

Both relative intensities and the vibrational isotope effect show that the bands in the region 2120–2360 A. belong to two different systems (figure 2). This is confirmed by the vibrational analysis. Most of the observed bands of the G←X system fit into what appears to be a $v'=0$ progression. The wave-number differences identify the absorbing molecule as SnTe. Possible values of ν_e and ω_e for the G level are given in table 7: they are subject to the same kinds of uncertainty as were discussed in connection with the F←X system of SnSe.

Table 6.   Band-head data for the G←X system of SnTe

Wave-length (I.A.)	ν, cm.$^{-1}$	$\Delta\nu$	Provisional v', v'' values
2117·5	47211		1,1
		227	
2127·7	46984		0,1
		257	
2139·4	46727		0,2
		256	
2151·2	46471		0,3
		260	
2163·3	46211		0,4
		259	
2175·5	45952		0,5
		258	
2187·8	45694		0,6

Table 7.   Summary of band-head constants

Molecule	State	ν_e	ω_e	$x_e\omega_e$	$y_e\omega_e$
SnSe	F	(47850)	(290)	—	—
	E	30738·9	196·6	−0·77	−0·0016
	D	27549·6	225·1	−0·69	—
	X	0	331·2	−0·736	—
SnTe	G	(47260)	(230)	—	—
	F	44033·5	229·7	−1·25	−0·003
	E	(28000)	(150)	—	—
	D	25444·3	179·1	−0·40	—
	X	0	259·5	−0·50	—

§ 4. DISCUSSION

It is interesting to compare the new results for SnSe and SnTe with those for SnO and SnS. The relevant values of ν_e, and of the force constant expressed as percentages of those for the ground states, are given in table 8. The ν_e values, both for the D and for the E states, decrease fairly regularly from oxide to telluride— as might have been expected. They suggest that the states labelled with a common letter may be spectroscopically similar, and this suggestion receives some support from the values of the force constants. Thus the D states have force constants which are all about 48% of those for the ground states, while the figures for the E levels are all close to 35%. This behaviour is quite analogous to that displayed by the D and E states of the corresponding silicon compounds (Vago and Barrow, 1946).

Table 8.   Values of ν_e, and of the force constants expressed as
percentages of those for the ground states

Molecule : State	SnO	SnS	SnSe	SnTe
G ·	—	—	—	(47260) (78%)
F	—	—	(47850) (77%)	44034 78·4% .
E	36803 · 36·6%	33037 36·4%	30739 35·1%	(28000) (33%)
D	29625 50·2%	28338 46·3%	27550 46·2%	25444 47·6%

The figures in table 8 also suggest that F and G levels of SnO and SnS and a G level of SnSe should exist. The force constants for these states may be about 78% of those in the ground states, so that ω_e'(SnO)~730, ω_e'(SnS)~430 cm.$^{-1}$ It is hardly possible to make any accurate prediction of the positions of the absorption systems involving these states, but they should lie in the vacuum. ultra-violet region within the rough limits 1600 to 2000 A.
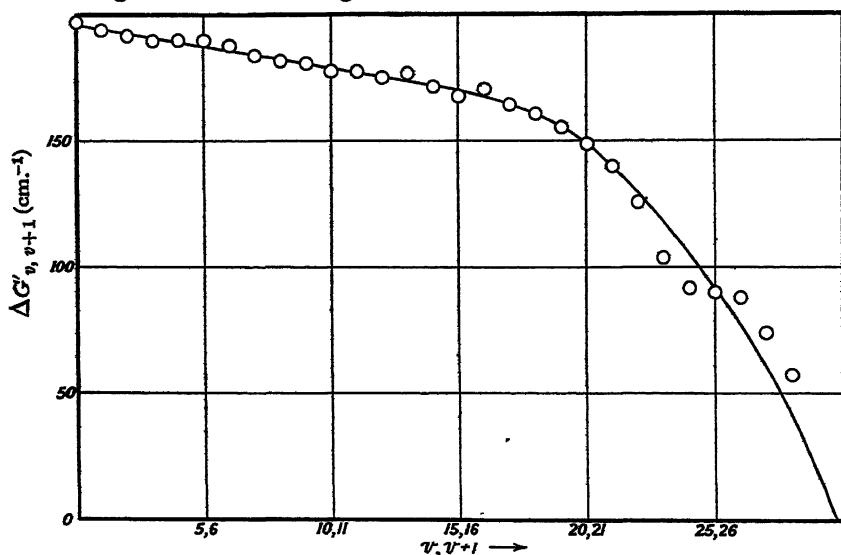
Figure 3.   The variation of vibrational interval with vibrational quantum number for the E state of SnSe.

It remains to consider the dissociation energies for these states. The most reliable figure is that for the E level of SnSe, where bands up to $v' = 29$ have been. observed (table 2): in this region the vibrational interval between successive levels is less than 30% of ω_e', so that a fairly good estimate of D_s should be obtained by integration of the curve got by plotting $\Delta G'_{v, v+1}$ against $v, v+1$. This plot is shown in figure 3: from it D_s is found to be 4800 cm.$^{-1}$. It is note- worthy that this figure is only 40% of that obtained by linear extrapolation of the first fifteen vibrational intervals.

The integrated value of D_s might be in error either by faulty location of the band-heads or by incorrect assignment of the v' values. The latter point has.

been discussed above: it is difficult to estimate the error arising from the uncertainty in the band-head measurements, for which the vibrational isotope effect is mainly responsible. For SnSe this effect is so complicated that we can hardly hope for detailed resolution of the various heads of a given band, while at the same time its magnitude is such that the total spread of a band distant 1000 cm.$^{-1}$ from the origin is nearly 25 cm.$^{-1}$ In these circumstances all that can be done is to measure throughout the strongest bands. This we have attempted to do, and it seems likely that the resulting limits of error of the integrated value of D_{E} are not very large. However, another figure for D_{E} may be derived by measuring the point at which it appears that the banded absorption passes over into a continuum. It is difficult to estimate the position of this point accurately, but it seems to lie at about 2730 A., giving $D_{\text{E}} = 5900$ cm.$^{-1}$, or 0·73 ev. The fact that this figure is a little higher than that got by integration of the curve in figure 3 suggests that the vibrational levels do not in fact converge quite so rapidly at high v-values as has been drawn. Nevertheless, the agreement between the two values of D_{E} may be regarded as fairly satisfactory, and we can be reasonably confident that the true value is close to the higher figure, 0·73 ev., that is, about 50% of the value obtained by linear extrapolation.

No such certain figure can be given for the E state of SnTe, but again the beginning of what appears to be a region of continuous absorption may be used to deduce the rough value 0·6 ev. Since, however, we have not been able to put forward a satisfactory vibrational analysis for the E–X system, this estimate of D_{E} must be treated with considerable reserve.

In order to obtain information about the dissociation energies in the ground and D states, it is necessary to consider the products of dissociation. For various reasons, some of which have been mentioned in a discussion of the dissociation energies of the X, D and E states of the silicon compounds (Vago and Barrow, 1946), we think that the most likely interpretation of the present data is that the X, D and E states each dissociate into a pair of ground state atoms, $^3P + {}^3P$, although other possibilities are not definitely excluded. Table 9 has been constructed on this assumption. If it is correct, then for any one molecule

$$D_{\text{X}} = \nu + D_{\text{D}} = \nu_{\text{E}} + D_{\text{E}}.$$

The values given above of D_{E} for SnSe and SnTe have been used: for SnO and SnS we have taken 50% (by analogy with SnSe) of the values obtained by linear extrapolation of the vibrational intervals. The values given for D_{D} are in all cases 70% of the linearly extrapolated values; these appear to us on general grounds to be reasonable, if somewhat arbitrary, estimates. For the ground states, for which the extrapolation is likely to be least in error, we have taken 100% of the linear values.

The quantities $\nu_{\text{E}} + D_{\text{E}}$, $\nu_{\text{D}} + D_{\text{D}}$ and D_{X} are given in rows v, vi, and vii of table 9. The agreement between them is fair—probably as good as can be expected. The least satisfactory value appears to be that of D_{D} for SnTe, which seems to be a good deal too high. The measure of concordance of the other values for a given molecule is some support for the view that the products of dissociation are the same in the three states, and leads us to attempt to derive improved values, first of D_{X} (for SnO and SnS the means of v, vi and vii have

been taken; for SnTe the mean of v and vii; and for SnSe the value of $D_\mathbf{x}$ determined from the beginning of continuous absorption has been accepted), and thence of D_D and $D_\mathbf{x}$. These improved values are given in rows viii, ix and x. As was found for the silicon compounds, there is a more or less regular decrease in the value of D for a given state on going from oxide to telluride, and of course in the values of D_x, D_D and $D_\mathbf{x}$ for a given molecule : the force constants follow similar general trends.

Table 9. Electronic energies and energies of dissociation (ev.)

		SnO	SnS	SnSe	SnTe
i	$\nu_\mathbf{x}$	4·56	4·10	3·81	3·4$_7$
ii	ν_D	3·67	3·51	3·42	3·15
iii	$D_\mathbf{x}$	1·50*	1·17*	0·73 †	(0·6) †
iv	D_D	2·39 ‡	1·91 ‡	1·59 ‡	1·74 ‡
v	$\nu_\mathbf{x}+D_\mathbf{x}$	6·06	5·27	4·54	(4·07)
vi	ν_D+D_D	6·06	5·42	5·01	4·89
vii	D_x	5·62 §	5·50 §	4·62 §	4·17 §
Improved values					
viii	D_x	5·91	5·40	4·54	4·1
ix	D_D	2·24	1·89	1·12	0·9
x	$D_\mathbf{a}$	1·35	1·30	0·73	0·6

* Obtained by taking 50% of the figures determined by linear extrapolation of the vibrational intervals.

† From an estimate of the onset of continuous absorption.

‡ obtained by taking 70% of the figures determined by linear extrapolation of the vibrational intervals.

§ Obtained by linear extrapolation of the vibrational intervals.

REFERENCES

BARROW, R. F., 1940. *Proc. Phys. Soc.*, **52**, 380.
BARROW, R. F., 1944. Discussion on Band Spectra. *Proc. Phys. Soc.*, **56**, 204.
BARROW, R. F., and VAGO, E. E., 1943. *Proc. Phys. Soc.*, **55**, 326 ; 1944. *Ibid.*, **56**, 78.
VAGO, E. E. and BARROW, R. F., 1946. *Proc. Phys. Soc.*, **58**, 538.

RADAR ECHOES FROM THE SEA SURFACE AT CENTIMETRE WAVE-LENGTHS

BY H. DAVIES AND G. G. MACFARLANE,
Telecommunications Research Establishment

ABSTRACT. An account is given of quantitative measurements of the echo obtained from the sea surface with radars operating at wave-lengths of 1·25, 3 and 10 cm. under various weather conditions from calm to very rough. The necessary theoretical treatment is given, and a function $f(\theta)$, called the *scattering coefficient* of the surface, is defined.

§1. INTRODUCTION

THE detection of small objects situated on the surface of the sea is limited by the fact that the sea in the vicinity of the target scatters backwards a certain amount of the incident energy, and the target echo is frequently masked by this scattered radiation. To effect any improvement in the discrimina-

-tion of target from " clutter ", a knowledge of the magnitude of sea clutter for
·different wave-lengths and various states of the sea, and its dependence on angle
·of depression and range, is essential. An understanding of the scattering
mechanism coupled with a theoretical treatment of the subject would be desirable.

A theoretical attack on the problems has been made,[*] in which it is assumed
that the surface of the sea consists of cylindrical sinusoidal waves, the wave-
length and wave-height of which both satisfy certain statistical distributions.
The model of the sea is a two-dimensional one.   However, the theoretical approach
is handicapped by inadequate knowledge of the exact wave-form of the sea waves
and the shape of the statistical distributions.   A mathematical treatment based
·on any wave-shape other than the simple sinusoid is extremely difficult, and if
·a solution is obtainable it is well nigh incapable of interpretation.

' In this paper emphasis is laid on the experimental results rather than on the
·theoretical problem.

§2. THEORETICAL CONSIDERATIONS

The notation given below is the one normally employed in radar, and will
be adopted throughout this article

Let P = Peak power output of transmitter.

G = Power gain of aerial system over an isotropic radiator.

A = Effective aperture of aerial system $\left(G = \frac{4\pi A}{\lambda^2}\right)$.

λ = Wave-length of radar.

τ = Pulse length (in time).

c = Velocity of E.M. waves in air.

h = Height of radar above the sea.

R = Slant range.

θ = Angle of depression from horizontal. ⎱
ϕ = Angle of azimuth from centre of beam. ⎰ See figure 1.

$F(\theta, \phi)$ = Amplitude polar diagram factor of aerial.

Φ = Beam width in azimuth, defined as the angular distance in radians
between directions 3 db. down on the maximum on either side
for one-way transmission.

Referring to figure 1, consider a small element of sea ds situated at P, at
range R, depression θ, and azimuth ϕ.

The power density at P is given by

$$S_i = \frac{PGF^2(\theta, \phi)}{4\pi R^2}. \qquad \dots\dots(1)$$

The power intercepted by the element ds is $S_i \sin \theta\, ds$. If this power were
scattered isotropically over the hemisphere above ds, then the power received
by the radar would be

$$dP_r = \frac{S_i \sin \theta\, ds\, A F^2(\theta, \phi)}{2\pi R^2} \qquad \dots\dots(2)$$

$$= \frac{PGA \sin \theta\, F^4(\theta, \phi)\, ds}{8\pi^2 R^4}. \qquad \dots\dots(3)$$

[*] Unpublished paper by G. G. Macfarlane.

This formula has been derived on the assumption that the element ds behaves as a " perfectly rough " surface, i.e. a surface which scatters all the incident energy uniformly over the hemisphere above it. This definition of a " perfectly rough " surface is contrary to the optical concept of such a surface, which scatters according to Lambert's cosine law, but it is more convenient as a reference level or standard surface.

In general it is not true to say that ds scatters isotropically, and we must write

$$dP_r = \frac{PGA \sin \theta F^4(\theta, \phi) f(\theta)\, ds}{8\pi^2 R^4}, \qquad \dots\dots (4)$$

where $f(\theta)$ is also a function of the radar wave-length and the state of the sea prevailing. $f(\theta)$ may be called the scattering coefficient, and is essentially a ratio.

The expression (4) is the power received back from the element of sea ds. By the very nature of the pulse technique employed in radar, echoes are received by the radar at the same time from all elements in an area which are at ranges lying between R and $R + \tau c/2$. This area is called the *resolvable area* of the radar. We assume that the surface behaves as an incoherent scatterer, and so dP_r must be summed or integrated over the resolvable area.
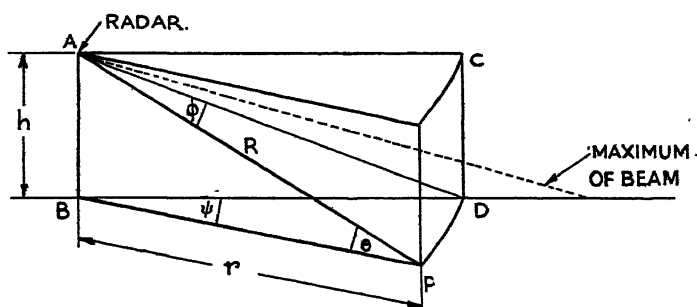
Figure 1. The configuration and the co-ordinates.
(*All figures Crown Copyright.*)

The resolvable area in this case for range R is the annulus between the two circles on the sea of slant radius R and $R + \tau c/2$. Assuming a symmetrical beam, most of the energy is concentrated in a cone of semi-vertical angle $\Phi/2$. It can be seen that we have essentially two different configurations with which to deal:

1. The case in which the resolvable area is narrow compared with the area illuminated by the main part of the beam. See figure 2.

2. The case in which the resolvable area is wide compared with the area illuminated by the main part of the beam. See figure 3.

We shall deal separately with these two cases.

§ 2.1.

The mathematical expression for this state of affairs is

$$\frac{\tau c}{2} \sec \theta \ll R\Phi \csc \theta.$$

The maximum return from the resolvable area at range R will be given when this area lies in the centre of the main part of the beam as shown in figure 2. In this case we may proceed thus. As the element of sea, we may take

$$ds = \frac{\tau c}{2} \sec \theta \cdot r \, d\psi \qquad \qquad \ldots\ldots(5)$$

$$= \frac{\tau c}{2} R \sec \theta \, d\phi.$$

$$\therefore \quad dP_r = \frac{PGA}{16\pi^2 R^4} R \tau c \tan \theta \cdot F^4(\theta, \phi) f(\theta) \, d\phi, \qquad \ldots\ldots(6)$$

$$P_r = \frac{PGA}{16\pi^2 R^3} \tau c \tan \theta f(\theta) \int F^4(\theta, \phi) \, d\phi. \qquad \ldots\ldots(7)$$

The area of integration is small enough to justify the removal of $R \tan \theta$ and $f(\theta)$ outside the integral sign. These quantities vary only slightly over the area concerned.
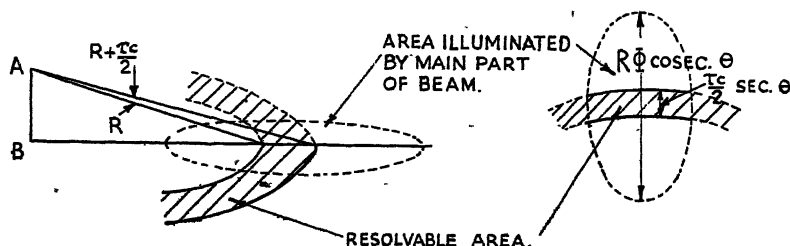
Figure 2. Resolvable area well within main area of illumination.

The integral in (7) may be evaluated from the polar diagram of the radar, but often a sufficiently good approximation is

$$\int F^4(\theta, \phi) \, d\phi = \Phi, \qquad \qquad \ldots\ldots(8)$$

$$P_r = \frac{PGA}{16\pi^2 R^3} \tau c \tan \theta f(\theta) \cdot \Phi. \qquad \ldots\ldots(9)$$

This result is valid only as long as $\frac{\tau c}{2} \sec \theta \ll R\Phi \operatorname{cosec} \theta$, which is so for sufficiently small angles of depression.

§ 2.2.

The mathematical condition for this state is

$$R\Phi \operatorname{cosec} \theta \ll \frac{\tau c}{2} \sec \theta.$$

We express ds as a function of both $d\theta$ and $d\phi$:

$$ds = R^2 \, d\theta \, d\phi \operatorname{cosec} \theta. \qquad \ldots\ldots(10)$$

$$\therefore \quad dP_r = \frac{PGA}{8\pi^2 R^2} f(\theta) F^4(\theta, \phi) \, d\theta \, d\phi. \qquad \ldots\ldots(11)$$

We again assume that $F^4(\theta, \phi)$ is a function of θ and ϕ which varies much more rapidly than R and $f(\theta)$.

$$\therefore \quad P_r = \frac{PGA}{8\pi^2 R^2} f(\theta) \int\int F^4(\theta, \phi) d\theta \, d\phi. \qquad \dots\dots(12)$$

This double integral is evaluated from a knowledge of the polar diagram, and is approximately constant as long as

$$R\Phi \cosec \theta \ll \frac{\tau c}{2} \sec \theta,$$

i.e. for sufficiently large angles of depression.

Thus it can be seen that for the echo power received from the surface there are two formulae, each applicable over a different region. If θ_0 is the value of θ for which $R\Phi \cosec \theta = \dfrac{\tau c}{2} \sec \theta$, then

> when $\theta \ll \theta_0$, equation (7) is valid;
>
> when $\theta \gg \theta_0$, equation (12) is valid.

In both cases $f(\theta)$ has appeared as a multiplying factor, and from the experimental results this important function may be easily calculated.
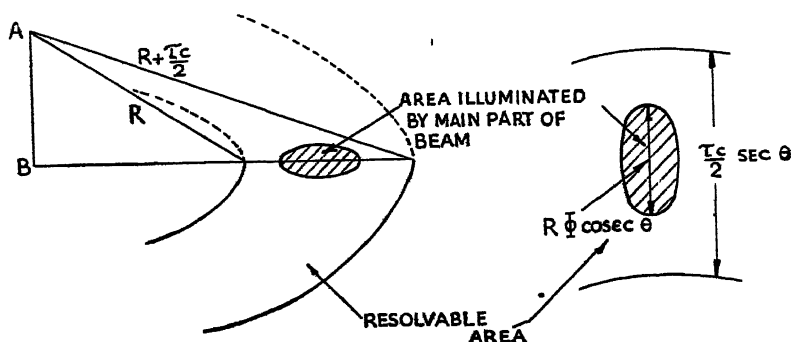
Figure 3. Main area of illumination well within resolvable area.

In the theoretical approach to the problem, the object was to determine mathematically the coefficient $f(\theta)$ in terms of θ, λ, and roughness of the sea. The starting point of the analysis was the assumption that a sea wave is sinusoidal in section. The model of the sea was two-dimensional, the waves being cylindrical.

The return from a single wave was first calculated, and then this echo averaged over all waves. The statistical distribution for the amplitudes of the waves was taken to be

$$p(y) = \frac{y}{A^2} \exp\left(\frac{-y^2}{A^2}\right) \qquad y > 0, \qquad \dots\dots(13)$$

where A is the mean amplitude.

Similarly the distribution assumed for the wave-lengths of the sea waves was

$$q(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\Delta)^2}{2\sigma^2}\right\}, \qquad \dots\dots(14)$$

where Δ is the mean wave-length and σ is the standard deviation.

On the basis of these assumptions it has been shown * that $f(\theta)$ should be independent of θ for angles greater than some critical angle, and should behave as θ^n for very small angles. We shall see later that these conclusions are not altogether borne out by the results.

We should not expect the mathematical treatment to give completely correct results, since besides the assumptions already mentioned, no account has been taken of "white horses", which undeniably destroy the sinusoidal shape of the sea waves.

§3. EXPERIMENTAL DETAILS

The results were obtained from three separate sets of experiments which will be dealt with in order of importance.

§3.1. Using four radar equipments, a systematic investigation into sea clutter was conducted by Telecommunications Research Establishment at Great Orme's Head, Llandudno, 500 ft. above sea-level, with a wide angle of view straight out to sea. Two of the radars, K3 and K4, operated on a wave-length of 1·25 cm., one, X3, on a wave-length of 3 cm., and one, S1, on a wave-length of ·10 cm. X3 had a conical beam of width 4°·5, a peak power output of 30 kw., and a pulse length of 0·5 μsec. K4 had a beam of width 0°·58 in azimuth and 15° in elevation, a peak power of 20 kw., and could send out pulses of either 0·5 μsec. or 0·2 μsec. duration.

Calibration of the radars was effected by injecting a measured voltage at the I.F. stage and observing the deflection on the cathode-ray tube. In order to measure the absolute value of $f(\theta)$, and not merely its variations, measurements were made on the echo from a metal sphere. It is well known that the echo power P_s from a sphere of radius a is given by

$$P_s = \frac{PG}{4\pi R^2} \frac{\pi a^2 A}{4\pi R^2} \qquad \text{if } a \gg \lambda. \qquad \ldots\ldots(16)$$

If we compare this with (9), we see that

$$\frac{P_r}{P_s} = \frac{R\tau c \tan\theta . f(\theta) . \Phi}{\pi a^2}$$

$$= \frac{h\tau c\Phi f(\theta)}{\pi a^2} . \qquad \ldots\ldots(17)$$

Thus $f(\theta)$ may be calculated without the measurement of the rather awkward quantities P, G, A, etc.

The procedure adopted for measuring the sea echo at range R was to tilt the beam until maximum signal was obtained at range R on the tube. In this case the range of angles investigated was $\frac{1}{2}° < \theta < 5°$, a region well within the condition $\theta < \theta_0$. Formulae (7) and (9) are applicable to all these results.

§3.2. The second set of experiments was conducted in an aircraft and the range of angles covered was $15° < \theta < 90°$. Only an X Band Set was used, with a beam width of 6°, a pulse length of 1 μsec. and a peak power output of 20 kw. Unfortunately no measurements on a sphere were made, and so only variations

* Unpublished paper by G. G. Macfarlane.

in $f(\theta)$ could be observed.   However, some idea of the absolute value of $f(\theta)$ can be obtained by linking up the results of this set with those for X Band at Llandudno.

§ 3.3.   A number of observations on S Band have been made by the Army Operational Research Group at Ashford Flank.   The range covered is small, being $1° < \theta < 5°$, when equation (7) is valid.   From these measurements $f(\theta)$ has been calculated.

§ 4. THE RESULTS

§ 4.1. *The shape of the echo*

If the resolvable area is very large, as is the case for a high-flying aircraft with a radar beam width of say 6° and a pulse length of 1 μsec., then the echo appearing on the cathode-ray tube looks like figure 4 (*a*).   If the resolvable area is rather smaller, then the appearance of the echo is as shown in figure 4 (*b*).   The latter case was observed at Llandudno.   Two levels were discernible which we shall call the peak and mean sea levels, the mean sea level being about 6 db. down on the peak sea.   If the resolvable area is made very small, then ultimately individual waves will be resolved and the echo will consist of a large number of separate " blips ".
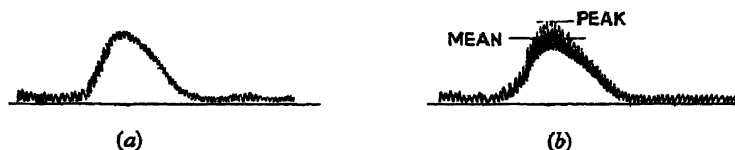
Figure 4.   The appearance of the echo on "A" scope.

The appearance of the echo

We have concentrated on reading the peak level when two levels have existed, and although it is not certain that this level corresponds to the P_r of § 2, there is reason to believe that it behaves in a similar manner.   Our calculation of $f(\theta)$ refers to this level.

§4.2. *Results on $\lambda = 1·25$ cm.*

Over the limited range for which results are available, i.e. $\frac{1}{2}° < \theta < 5°$, the results have shown that the echo power P_r varies inversely as the fourth power of the range.   Figure 5 is an example, typical of many such observations.

Since $h = R \operatorname{cosec} \theta \approx R \cot \theta$,

$$P_r \propto \frac{1}{R^3} \tan \theta,$$

from which it is clear on referring to equation (9) that $f(\theta)$ is independent of θ in the interval $\frac{1}{2}° < \theta < 5°$.

Although $f(\theta)$ is independent of θ over the interval $\frac{1}{2}° < \theta < 5°$, it is not independent of the state of the sea.   As a criterion of the degree of roughness of the sea, it was thought best to make observations on the wave-height.   This was done from a trawler and a submarine operating in the vicinity.   The values of $f(\theta)$ for different seas are given in figure 6, from which it can be seen that $f(\theta)$ does not increase perceptibly after a wave-height of 3 feet.   Even though no detailed readings are available for wave-heights of the order of 6 to 12 inches, it is known

that the echo decreases very rapidly in this region. The maximum value of $10 \log f(\theta)$ observed on $\lambda = 1 \cdot 25$ cm. was -12.

On the equipment K4 it was possible to verify the reasoning which leads to the echo power being proportional to the pulse length. Decreasing τ from
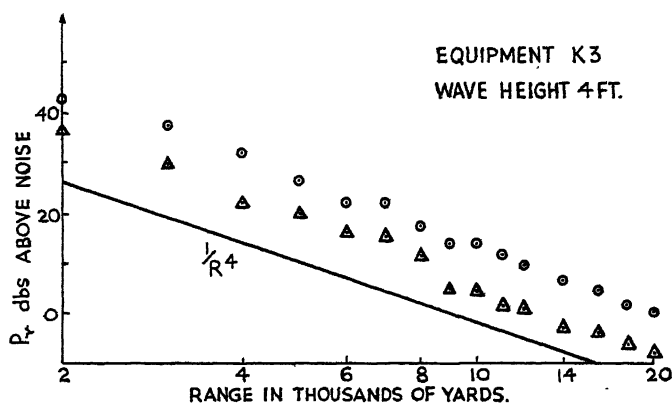
Figure 5. The dependence of echo power on range for $\lambda = 1 \cdot 25$ cm.

$0 \cdot 5 \, \mu\text{sec.}$ to $0 \cdot 2 \, \mu\text{sec.}$ produced the predicted drop in echo power to within the limits of experimental error.

§ 4.3. *Results on wave-length of 3 cm.*

As indicated previously, there are two sources of information on $\lambda = 3$ cm. The range $\frac{1}{2}° < \theta < 5°$ is covered by the Llandudno experiments, with which we shall deal first.
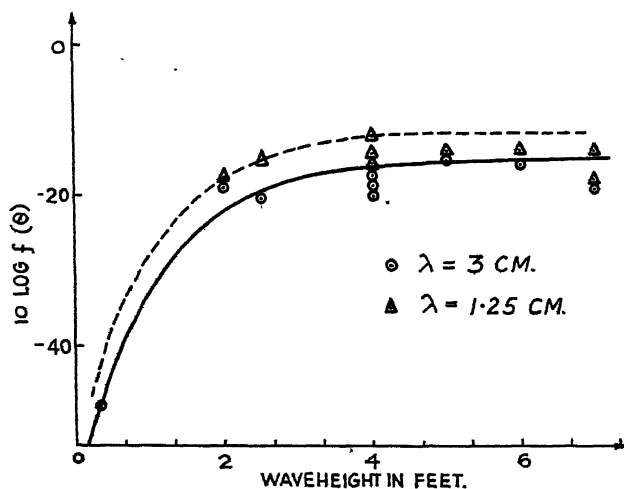
Figure 6. The dependence of $f(\theta)$ on the state of the sea.

The results are very similar to those on $1 \cdot 25$ cm. wave-length. Over the interval $\frac{1}{2}° < \theta < 5°$ the echo was found to follow the $1/R^4$ law, implying that $f(\theta)$ is independent of θ. A very large number of results confirms this; a typical example is shown in figure 7.

For the dependence of $f(\theta)$ on the state of the sea the reader should consult figure 6, where $10 \log f(\theta)$ over the interval $\frac{1}{2}° < \theta < 5°$ is plotted against wave-height. Three feet appears to be the height above which $f(\theta)$ does not vary with
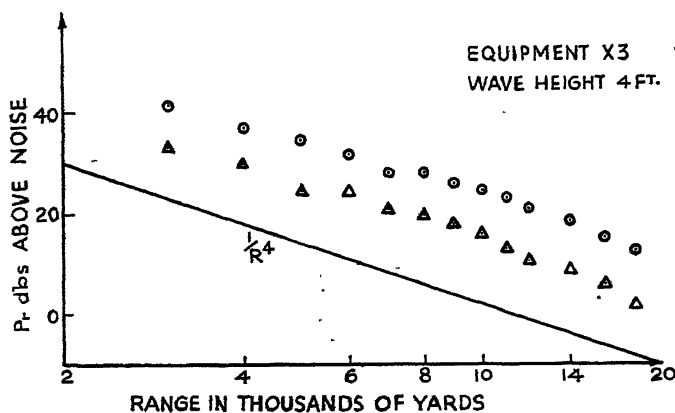
Figure 7. The dependence of echo power on range for $\lambda = 3$ cm.

the height of the waves. The maximum observed value of $10 \log f(\theta)$ was -15.5.

From the second source of information on 3 cm. wave-length, the aircraft results, we have been able to calculate $f(\theta)$ over the whole range $15° < \theta < 90°$. The variation with θ for one particular state of the sea is shown in figure 8.

The constancy of $f(\theta)$ found at Llandudno appears to extend up to $40°$. As before, the level of AB, the flat portion of the curve, depends on the roughness of the sea, but no wave-height measurements are available and so there is no
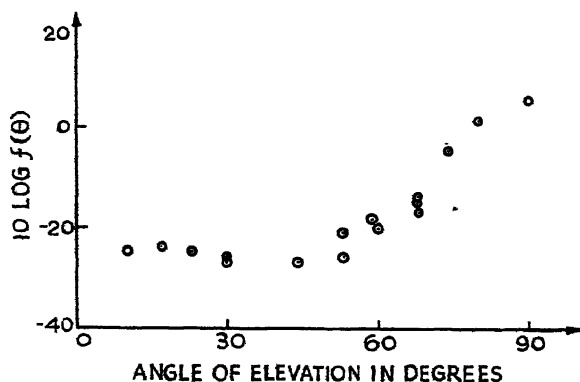
Figure 8. The dependence of $f(\theta)$ on θ, the angle of elevation ($\lambda = 3$ cm.).

check on the data in figure 6. The aircraft results indicate that, although the level AB depends on the state of the sea, $10 \log f(90°)$ is independent of this variable and the value of $10 \log f(90°)$ is $+5$. This means that the echo from the sea at normal incidence is 5 db. greater than the echo obtained from a perfectly rough surface in the same direction.

Let us compare the echo received back from a perfectly smooth conducting surface with the echo obtained from the sea surface, both being viewed at normal incidence on 3 cm.

For the sea surface,

$$P_r = \frac{PGA}{8\pi^2 R^2} f(\theta) \iint F^4(\theta, \phi)\, d\theta \cdot d\phi,$$

where $10 \log f(\theta) = 5$; or we may avoid the double integral by using the following reasoning, which is less accurate.

All the power incident on the sea falls within the resolvable area.

∴ The power intercepted is P.

∴ Power received back is given by

$$P_r = \frac{P \cdot f(\theta) G \lambda^2}{2\pi R^2 4\pi}$$

$$= \frac{PG\lambda^2 f(\theta)}{8\pi^2 R^2}.$$
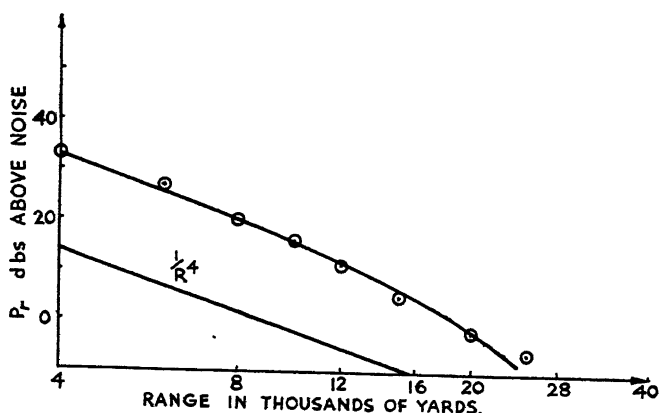
Figure 9. The dependence of echo power on range for $\lambda = 10$ cm.

The power P_r' received from a perfectly smooth surface is the power received by the radar if it were situated at the image point

$$P_r' = \frac{PGA}{4\pi(2R)^2} = \frac{PG^2\lambda^2}{64\pi^2 R^2}.$$

$$\therefore \quad \frac{P_r}{R_r'} = \frac{8f(\theta)}{G} \approx \frac{25}{G}.$$

Therefore, if the aerial gain is greater than 25, more energy is received back from a perfectly smooth surface than from the sea.

§ 4.4. *Results on wave-length of 10 cm.*

The relatively few results on 10 cm. wave-length indicate that the echo power varies inversely as the fourth power of the range for angles of depression from 5° down to 1°. Thereafter the decrease is more rapid. This is clearly shown in figure 9.

Thus, for the interval $1° < \theta < 5°$, $f(\theta)$ is independent of θ. Its dependence upon the state of the sea in this region of angles is shown in figure 10. Unfortunately, measurements of wave-height were not made, and it is doubtful whether the use of wind-speed as a criterion of the degree of roughness is justifiable. Some attempt has been made to replace wind-speed by wave-height. The high degree of scatter in the results is undoubtedly due to this bad criterion.

It appears that on 10 cm. wave-length $f(\theta)$ behaves somewhat the same as did $f(\theta)$ on 3 cm. and 1·25 cm., but the maximum value is only reached at a wave-height considerably greater than 3 ft. This maximum value of $10 \log f(\theta)$ is -21.
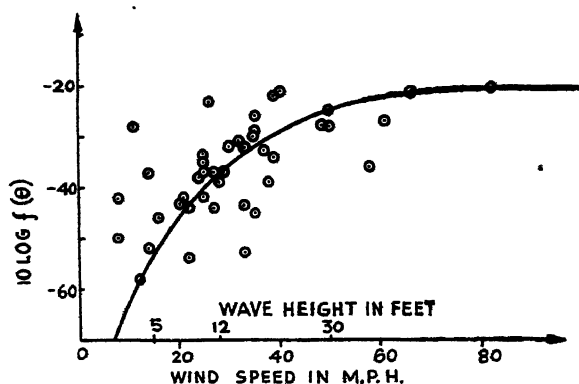
Figure 10. The dependence of $f(\theta)$ on the wind velocity ($\lambda = 10$ cm.).

§ 4.5. *The dependence of $f(\theta)$ on λ*

We have seen that over a considerable range of angles $f(\theta)$ is independent of θ, but dependent on the state of the sea. Its value increases as the sea gets rougher, tending to a maximum value which depends on the wave-length of the radar. This dependence is clear from the accompanying table.

Wave-length	Max. value of $10 \log f(\theta)$
1·25 cm.	-12
3 cm.	$-15·5$
10 cm.	-21

This leads immediately to the result

$$f(\theta)_{max} = \frac{\lambda_0}{\lambda},$$

where $\lambda_0 = 1/12$ cm. and λ is expressed in cm.

It might be thought that $f(\theta)$ would tend to the same value on all wave-lengths as the wave-heights increase, but this is obviously not the case, anyway not for small angles. This fact suggests that the greatest echo is obtained from ripples and not from the large waves, and that although the wave-height increases to very large values, the quantity and size of ripple do not increase after a certain wave-height has been achieved.

§5. CONCLUSIONS

Although the results so far obtained are not inconsiderable, there is much room for expansion in the subject. Only on 3 cm. wave-length has the whole range of angles of depression been investigated. The same information on K and S bands would not be without advantage. Then the scattering coefficient $f(\theta, \lambda,$ state of sea) would be known for all θ, for $\lambda = 1$ to 10 cm., and for all seas from calm to very rough. When this coefficient is known, the echo obtained from the sea with any radar equipment may be calculated.

Another possible extension is the examination of the shape of the scattering polar diagram. Up till now, only the echo in the direction of incidence, i.e. the radar echo, has been investigated. With a receiver completely detached from the transmitter, it would be possible to investigate the energy scattered in other directions, and so to determine the scattering polar diagram. It would be interesting and informative to determine the scattering polar diagram for a number of angles of incidence. The reflection coefficient of the surface might be measured at the same time.
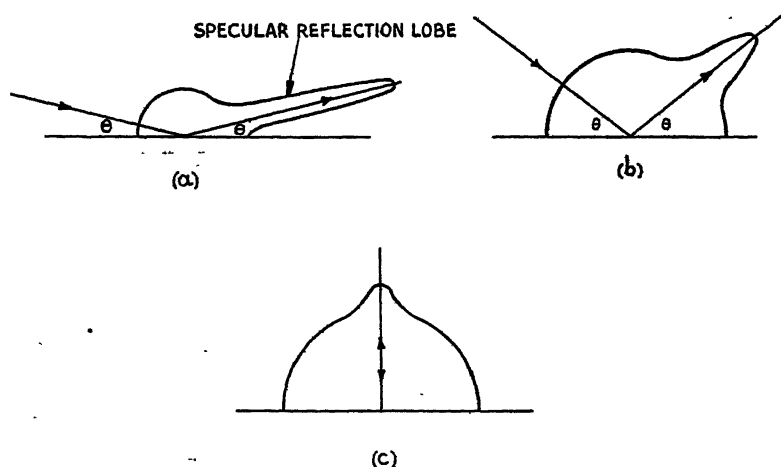
Figure 11. Hypothetical scattering polar diagram.

It is anticipated that the section of the scattering polar diagram in the plane of incidence will be similar to those shown in figure 11. For small angles of incidence the reflection lobe will be very large compared with the scattered energy. As the angle increases, the reflection lobe moves as shown and also becomes smaller. Thus more energy is scattered than in figure 11 (a). Finally, at normal incidence, the lobe of specular reflection is barely significant, the return being only 5 db. above that from a " perfectly rough " surface. However, the diagrams given are purely hypothetical, and experiments will have to be made before any reliance at all can be placed upon them.

§6. ACKNOWLEDGMENTS

of the Ministry of Aircraft Production, and is now published by permission of the Ministry of Supply (Air). All the diagrams are Crown Copyright.

Thanks are due to J. R. Atkinson, A. L. Cockcroft, E. R. Wiblin, and G. J. R. MacLusky, who were responsible for the experimental work; to R. A. Finlayson, who took part in the earlier analysis of the results; to H. G. Hooker for helpful criticism throughout; to the Superintendent, Army Operational Research Group, for certain results at $\lambda = 10$ cm.; and finally, to the Royal Navy for its helpful co-operation during the experiments.

EXTENSION OF GRIFFITH'S THEORY OF RUPTURE TO THREE DIMENSIONS

By R. A. SACK,

H. H. Wills Physical Laboratory, University of Bristol

(Now at Department of Applied Mathematics, University of Liverpool)

ABSTRACT. Griffith's theory of rupture of brittle materials is extended to materials containing circular cracks. It is found that (*a*) the tensile strength of brittle material in one direction is not affected by stresses at right angles to it ; (*b*) the result differs from Griffith's by a factor depending on Poisson's ratio of the material and lying between 1·57 and 1·81.

§ 1. INTRODUCTION

A THEORY of rupture of brittle solids has been developed by Griffith (1921). This theory assumes the existence of a large number of small cracks on the surface or in the interior of the solid, rupture being primarily conditioned by the extension of a crack already existing, and not by the formation of a new one. The spreading of a crack is accompanied by an increase in energy proportional to the increase of the surface; on the other hand, the internal surface increases the elastic deformability of the material, thus leading to a lower free energy when the material is subject to given external stresses. It depends on the magnitudes and directions of crack and stresses whether the total contribution is positive or negative; if an increase of the size of a crack leads to a diminution of the free energy, the system will become unstable, the crack will spread and the material be fractured.

The quantitative development of Griffith's theory has to take account of the size and shape of the crack and of its orientation relative to the applied stress. For mathematical reasons, Griffith (1921, 1924) has confined the exact theoretical treatment to the problem in two dimensions, i.e. that of a plane plate with an elliptic cavity going right through it. By considering the crack as an ellipse of vanishing minor axis and assuming that Hooke's law holds good up to the corners of the crack, he finds that rupture will occur when the stress P normal to a crack

reaches the critical value

$$P_c = \sqrt{\left(\frac{2ET}{\pi c}\right)} \qquad \text{for plane stress,} \qquad \ldots\ldots(1)$$

$$P_c = \sqrt{\left(\frac{2ET}{\pi c(1-\sigma^2)}\right)} \quad \text{for plane strain,*} \quad \ldots\ldots(2)$$

where E denotes Young's modulus of the material, σ Poisson's ratio, T the surface tension, and c the width of the crack if situated on the surface of the plate or half this width if situated in the interior.

The origin of the Griffith cracks has not been properly explained, but for some crystals their existence has been demonstrated experimentally (Andrade and Martindale, 1936). If they are situated on the surface, their length is very likely to be much greater than their width c, so that the two-dimensional model is a good approximation. Glass and many crystals are materials of this type; their tensile strength can be varied considerably by suitable treatment of their surface (Joffe and Levitzky, 1925; Reinkober, 1937). If, however, the cracks are situated in the interior of the material, there is no reason to expect that their length will greatly exceed their width, and a three-dimensional model has to be used.

It is the purpose of this paper to provide the required three-dimensional theory. The conditions of rupture are calculated for a brittle amorphous solid containing a number of plane circular cracks when one of the principal stresses acts normally to the plane of one of these cracks. This normal principal stress must be tensile, for if it is compressive the faces of the crack will be pressed together and can exert tractions on each other, whereas the calculations are based on the assumption that the internal surface is free from tractions. The influence of other cracks can be neglected as the additional stresses and strains set up by them fall off rapidly with distance.

In § 2 the strain correction is calculated, which is set up by the crack if the material is subject to a homogeneous stress †; in § 3 the diminution of the free energy of the system due to the crack is determined, and the tensile strength is calculated in § 4. The case of oblique fracture connected with compressive external stresses will not be considered in this paper.

Since the present calculations were first carried out, I. N. Sneddon (1945) has calculated independently the stresses set up by a circular crack subject to an internal pressure. Instead of the curvilinear coordinates employed in the present paper, Sneddon uses cylindrical polars; he arrives at the displacements and stresses by the method of Hankel transforms developed by Harding and himself (1945). The advantage of Sneddon's method is that the stresses are obtained in well-determined directions; against this must be set the drawback that the evaluation of the stresses involves the use of rather lengthy integrals. The final result for the tensile strength is the same for both methods.

§2. CALCULATION OF STRAIN CORRECTION

To calculate the strain correction set up by the crack a method due to Neuber (1934) will be used. If $x_i(i=1, 2, 3)$ are the Cartesian coordinates of a point

---

* The formula $\sqrt{[2ET(1-\sigma^2)/\pi c]}$ for plane strain as stated by Griffith (1924) is obviously erroneous.

† This calculation forms a special case of a problem treated by Neuber in his dissertation. (Munich, 1932) ; this, however, has not been available to the writer.

within the elastic solid and q_i the displacements from the unstrained state in the direction of the three axes, Hooke's law can be expressed by the equation

$$S_{ik} = G\left[\frac{\partial q_i}{\partial x_k} + \frac{\partial q_k}{\partial x_i} + \delta_{ik}\frac{2\sigma}{1-2\sigma}\operatorname{div}q\right], \quad i,\,k = 1,\,2,\,3, \quad \ldots\ldots(3)$$

where the S_{ik} denote the stress components in rectangular coordinates, G is the modulus of rigidity, and

$$\delta_{ik} = \begin{cases} 1 \\ 0 \end{cases} \text{ if } \begin{matrix} i=k \\ i\neq k \end{matrix},$$

$$\operatorname{div}q = \frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} + \frac{\partial q_3}{\partial x_3}. \qquad \ldots\ldots(4)$$

If no external volume forces are acting, the condition for equilibrium is

$$\sum_k \frac{\partial S_{ik}}{\partial x_k} = 0, \qquad i=1,\,2,\,3, \qquad \ldots\ldots(5)$$

which with (3) leads to

$$\Delta q_i + \frac{1}{1-2\sigma}\frac{\partial}{\partial x_i}\operatorname{div}q = 0, \qquad i=1,\,2,\,3. \qquad \ldots\ldots(6)$$

According to Neuber the general solution of (6) is given by

$$q_i = -\frac{\partial F}{\partial x_i} + 2\alpha V_i, \qquad i=1,\,2,\,3, \qquad \ldots\ldots(7)$$

where

$$F = V_0 + x_1 V_1 + x_2 V_2 + x_3 V_3, \qquad \ldots\ldots(8)$$
$$\alpha = 2(1-\sigma) \qquad \ldots\ldots(9)$$

and the V_k are harmonic functions satisfying

$$\Delta V_k = 0, \qquad k=0,\,1,\,2,\,3. \qquad \ldots\ldots(10)$$

In the expressions (7) and (8) one of the functions $V_k(k\neq 0)$ can be put equal to zero without loss of generality.

If, instead of Cartesians, curvilinear orthogonal coordinates u_ν are used, such that an element of length ds at any point is given by

$$ds^2 = \sum_\nu h_\nu{}^2 du_\nu{}^2 \qquad \ldots\ldots(11)$$

(Greek suffixes λ, μ, ν refer to curvilinear coordinates), where the h_ν are functions of the u_ν, then the displacements q_ν have to be taken in the direction of increasing u_ν and the stresses $S_{\mu\nu}$ in the corresponding planes. The relation (3) has now to be replaced by

$$\frac{S_{\mu\nu}}{G} = \frac{h_\mu}{h_\nu}\frac{\partial}{\partial u_\nu}\left(\frac{q_\mu}{h_\mu}\right) + \frac{h_\nu}{h_\mu}\frac{\partial}{\partial u_\mu}\left(\frac{q_\nu}{h_\nu}\right) + \delta_{\mu\nu}\left[\frac{2-\alpha}{\alpha-1}\operatorname{div}q + \frac{2}{h_\mu}(q\cdot\operatorname{grad}h_\nu)\right], \quad \ldots\ldots(12)$$

where

$$\operatorname{div}q = \frac{1}{h_\lambda h^\mu h}\left[\frac{\partial}{\partial u_\lambda}(h_\mu h_\nu q_\lambda) + \ldots\right], \qquad \ldots\ldots(13)$$

whereas the Laplacian Δ is expressed by

$$\Delta = \frac{1}{h_\lambda h_\mu h_\nu}\left[\frac{\partial}{\partial u}\left(\frac{h_\mu h_\nu}{h_\lambda}\frac{\partial}{\partial u_\lambda}\right) + \ldots\right]. \qquad \ldots\ldots(14)$$

(The dots denote cyclic permutation of the suffixes λ, μ, ν.)

Though the functions V_k can be expressed in terms of the new coordinates u_ν, it is not as a rule possible to take them along directions that vary from point to point; they still have to be taken along the Cartesian axes. As

$$\cos(x_k, u_\nu) = \frac{1}{h_\nu} \frac{\partial x_k}{\partial u_\nu}, \qquad \ldots\ldots(15)$$

the solution (7) becomes

$$q_\nu = \frac{1}{h_\nu}\left(-\frac{\partial F}{\partial u_\nu} + 2\alpha \sum_k \frac{\partial x_k}{\partial u_\nu} V_k\right), \qquad \ldots\ldots(16)$$

where F is still defined by (8).

For problems relating to a plane circular disc or crack of infinitely small thickness, the appropriate coordinates are oblate spheroidal coordinates ψ, ϑ, ϕ, where the surfaces

$\psi = $ const. represent oblate spheroids,

$\vartheta = $ const.   ,,   hyperboloids of revolution of one sheet,

$\phi = $ const.   ,,   planes through the x_1-axis.

The Cartesians are expressed in the new coordinates as

$$\left.\begin{array}{l} x_1 = c \sinh \psi \cos \vartheta, \\ x_2 = c \cosh \psi \sin \vartheta \cos \phi, \\ x_3 = c \cosh \psi \sin \vartheta \sin \phi. \end{array}\right\} \qquad \ldots\ldots(17)$$

The surface $\psi = 0$ corresponds to the circular disc $x_1 = 0$, $x_2^2 + x_3^2 \leqslant c^2$; each point of the circle is represented twice, corresponding to the two sides of the disc. The functions h_ν are derived from (17),

$$\left.\begin{array}{l} h_\psi^2 = h_\vartheta^2 = c^2(\sinh^2 \psi + \cos^2 \vartheta) = c^2 h^2, \\ h_\phi = c \cosh \psi \sin \vartheta. \end{array}\right\} \qquad \ldots\ldots(18)$$

Thus the Laplacian becomes

$$\Delta = \frac{1}{c^2 h^2 \cosh \psi \sin \vartheta}\left[\frac{\partial}{\partial \psi} \cosh \psi \sin \vartheta \frac{\partial}{\partial \psi} + \frac{\partial}{\partial \vartheta} \cosh \psi \sin \vartheta \frac{\partial}{\partial \vartheta}\right]$$
$$+ \frac{1}{c^2 \cosh^2 \psi \sin^2 \vartheta} \frac{\partial^2}{\partial \phi^2}. \qquad \ldots\ldots(19)$$

The normal solutions of the equation $\Delta V = 0$ are

$$\Pi_n^m(\sinh \psi) P_n^m(\cos \vartheta) e^{im\phi}$$

and

$$\left.\begin{array}{l} \\ T_n^m(\sinh \psi) P_n^m(\cos \vartheta) e^{im\phi}, \end{array}\right\} \qquad \ldots\ldots(20)$$

where the functions P_n^m are the usual Legendre polynomials, whereas $\Pi_n^m(x)$ and $T_n^m(x)$ represent the Legendre functions of the first and second kind respectively of the argument ix (cf. e.g. Hobson, 1931), or explicitly

$$\left.\begin{array}{l} P_n^m(x) = \text{const.} \cdot (1-x^2)^{m/2} \dfrac{d^{n+m}}{dx^{n+m}}(1-x^2)^n, \\[2mm] \Pi_n^m(x) = \text{const.} \cdot (1+x^2)^{m/2} \dfrac{d^{n+m}}{dx^{n+m}}(1+x^2)^n, \\[2mm] T_n^m(x) = \text{const.} \cdot \Pi_n^m(x) \displaystyle\int_x^\infty \dfrac{dx}{[\Pi_n^m(x)]^2(1+x^2)}. \end{array}\right\} \qquad \ldots\ldots(21)$$

If one of the principal stresses P of the external system is normal to the crack, there will be no shear stresses $S_{\psi\vartheta}$ and $S_{\psi\phi}$ acting across its faces, neither will the other two principal stresses be affected by the presence of the crack, but the normal component $S_{11} = S_{\psi\psi} = P$ must be compensated for as there should be no tractions exerted on the inner surface. This disturbance will exhibit axial symmetry and fall off with increasing distance. Thus the only terms to be considered are those containing $T_n{}^0$; terms with $m \neq 0$ or $\Pi_n{}^0$ need not be considered in the calculations.

The first three functions T are

$$
\left.
\begin{aligned}
T_0(\sinh \psi) &= T_0{}^0(\sinh \psi) = \cot^{-1}(\sinh \psi), \\
T_1(\sinh \psi) &= T_1{}^0(\sinh \psi) = T_0(\sinh \psi)\sinh \psi - 1, \\
T_2(\sinh \psi) &= T_2{}^0(\sinh \psi) = T_0(\sinh \psi)(3 \sinh^2 \psi + 1) - 3 \sinh \psi.
\end{aligned}
\right\} \quad \ldots\ldots(22)
$$

If in (7), (8) and (16) we put $V_2 = V_3 = 0$ the displacements q_ψ and q_ϑ are

$$
\left.
\begin{aligned}
q_\psi &= \frac{1}{h}\left[-\frac{1}{c}\frac{\partial V_0}{\partial \psi} + (2\alpha - 1)\cosh \psi \cos \vartheta V_1 - \sinh \psi \cos \vartheta \frac{\partial V_1}{\partial \psi} \right], \\
q_\vartheta &= \frac{1}{h}\left[-\frac{1}{c}\frac{\partial V_0}{\partial \vartheta} - (2\alpha - 1)\sinh \psi \sin \vartheta V_1 - \sinh \psi \cos \vartheta \frac{\partial V_1}{\partial \vartheta} \right].
\end{aligned}
\right\} \quad \ldots\ldots(23)
$$

According to (12) the stresses acting across the faces $\psi = 0$ are given by

$$
\left.
\begin{aligned}
\frac{S_{\psi\psi}}{2G}\bigg|_{\psi=0} &= \frac{1}{c^2 \cos^2 \vartheta}\left[-\frac{\partial^2 V_0}{\partial \psi^2} + \tan \vartheta \frac{\partial V_0}{\partial \vartheta} + \alpha c \cos \vartheta \frac{\partial V_1}{\partial \psi} \right]_{\psi=0}, \\
\frac{S_{\psi\vartheta}}{2G}\bigg|_{\psi=0} &= \frac{1}{c^2 \cos^2 \vartheta}\left[-\frac{\partial^2 V_0}{\partial \psi \partial \vartheta} - \tan \vartheta \frac{\partial V_0}{\partial \psi} + (\alpha - 1)c \cos \vartheta \frac{\partial V_1}{\partial \vartheta} \right]_{\psi=0}.
\end{aligned}
\right\} \quad \ldots\ldots(24)
$$

The required conditions

$$
S_{\psi\vartheta} = 0, \quad S_{\psi\psi} = \text{const.} \quad \text{for } \psi = 0 \qquad \ldots\ldots(25)
$$

are satisfied by a disturbance specified by the harmonic functions V_0 and V_1:

$$
\left.
\begin{aligned}
V_0 &= \tfrac{1}{3}(\alpha - 1)c^2 A[T_0(\sinh \psi) + \tfrac{1}{2}T_2(\sinh \psi)(\tfrac{3}{2}\cos^2 \vartheta - \tfrac{1}{2})], \\
V_1 &= cAT_1(\sinh \psi) \cos \vartheta.
\end{aligned}
\right\} \quad \ldots\ldots(26)
$$

Introduction of these functions into (24) yields

$$
S_{\psi\vartheta} = 0, \quad S_{\psi\psi} = \pi GA, \quad \text{for } \psi = 0. \qquad \ldots\ldots(27)
$$

At large distances the spheroidal coordinates approximate spherical polars with

$$
c \sinh \psi = r \qquad \ldots\ldots(28)
$$

and ϑ and ϕ in their usual meanings. The functions T become

$$
T_0(\sinh \psi) = \frac{c}{r} - \frac{c^3}{3r^3} + \ldots, \quad T_1(\sinh \psi) = -\frac{c^2}{3r^2} + \ldots, \qquad \ldots\ldots(29)
$$

whereas T_2 can be neglected as it falls off in a higher power. The functions h_ν are now

$$
h_r = 1, \quad h_\vartheta = r, \quad h_\phi = r \sin \vartheta, \qquad \ldots\ldots(30)
$$

so that according to (12) the stresses are now expressed as

$$
S_{rr} = G\left[2\frac{\partial q_r}{\partial r} + \frac{2-\alpha}{\alpha-1}\operatorname{div} q \right], \quad S_{r\vartheta} = G\left[\frac{\partial q_\vartheta}{\partial r} - \frac{q_\vartheta}{r} + \frac{1}{r}\frac{\partial q_r}{\partial \vartheta} \right]. \qquad \ldots\ldots(31)
$$

The deformations at large distances will be, according to (16), (26), (29) and (30),

$$q_r = \frac{c^3 A}{9r^2}[-(4-\alpha)-(4\alpha+2)(\tfrac{3}{2}\cos^2\vartheta-\tfrac{1}{2})]+\dots,$$

$$q = \frac{2c^3 A}{3r^2}(\alpha-1)\cos\vartheta\sin\vartheta+\dots, \qquad \dots\dots(32)$$

from which the stresses are derived by means of (31):

$$\frac{S_{rr}}{2G} = \frac{2Ac^3}{9r^3}[4-\alpha+(8+\alpha)(\tfrac{3}{2}\cos^2\vartheta-\tfrac{1}{2})]+\dots,$$

$$\frac{S_{r\vartheta}}{2G} = \frac{Ac^3}{3r^3}(4-\alpha)\sin\vartheta\cos\vartheta+\dots \qquad \dots\dots(33)$$

The dots in (29), (32) and (33) denote terms to a higher power in c/r, which can be neglected.

§3. ENERGY CORRECTION

After determining the additional stresses and deformations set up by the crack, we have to consider how the elastic energy of the stressed material is affected by such a disturbance. If the material has the shape of a large sphere of radius R_0 containing a circular crack, it would not be correct simply to take the solution for the uncracked sphere and to superpose on this the strains (23), (26), (32). The condition of equilibrium demands that the tractions across the external surface given by (12) should agree accurately with the imposed stresses. This condition can be satisfied by putting

$$q_r = \frac{c^3 A}{9r^2}[-(4-\alpha)-(4\alpha+2)P_2]+\frac{2(\alpha-1)}{4-\alpha}Br+CrP_2-(2-\alpha)Dr^3P_2,$$

$$q_\vartheta = \left[\frac{2c^3 A}{3r^2}(\alpha-1)-\tfrac{3}{2}Cr+(\alpha+\tfrac{3}{2})Dr^3\right]\sin\vartheta\cos\vartheta,$$

$$\dots\dots(34)$$

where

$$P_2 = P_2^0(\cos\vartheta) = \tfrac{3}{2}\cos^2\vartheta-\tfrac{1}{2}, \qquad \dots\dots(35)$$

and by virtue of (31)

$$\frac{S_{rr}}{2G} = \frac{2Ac^3}{9r^3}[(4-\alpha)+(8+\alpha)P_2]+B+CP_2+\left(1-\frac{\alpha}{2}\right)Dr^2P_2,$$

$$\frac{S_{r\vartheta}}{2G} = \left[\frac{4-\alpha}{3}\frac{c^3 A}{r^3}-\tfrac{3}{2}C+\frac{2-\alpha}{2}Dr^2\right]\sin\vartheta\cos\vartheta. \qquad \dots\dots(36)$$

The B term corresponds to a pure hydrostatic pressure, the term with C to a combination of pure shearing stresses; the D term is a correction with the property

$$\text{div } q = 7(\alpha-1)Dr^2P_2. \qquad \dots\dots(37)$$

At the crack only terms to zeroth power in c/R will have to be considered, the D term will be negligible and the vanishing of the tractions across the faces is expressed as

$$S_{11} = S_{\psi\psi} = 2G(\tfrac{1}{2}\pi A+B+C) = 0. \qquad \dots\dots(38)$$

The free energy of an elastic solid under the influence of stresses is equal to half the integral of the product of the surface tractions into the displacements; this integral has to be taken over the whole surface and to be given a negative sign (cf. Griffith, 1921).

In the present case, in view of the axial symmetry, this integral will be

$$-2W = 2\pi \int q_r S_{rr} R_0^2 \sin \vartheta d\vartheta + 2\pi \int q_\vartheta S_{r\vartheta} R_0^2 \sin \vartheta d\vartheta. \qquad \ldots\ldots(39)$$

As R_0 tends towards infinity, terms of W depending on a negative power of R_0 will vanish. This is the reason why in (32) and (34) higher powers in $1/R_0$ than the second have been neglected, and in (33) and (36) higher powers than the third.

If a pure hydrostatic pressure N is exerted, (36) and (38) lead to

$$\left.\begin{array}{c} \dfrac{N}{2G} = B + \tfrac{2}{9}(4-\alpha)\dfrac{c^3}{R_0^3} A, \\[2mm] 0 = \tfrac{1}{2}\pi A + B. \end{array}\right\} \qquad \ldots\ldots(40)$$

Hence

$$\left.\begin{array}{c} B = \dfrac{N}{2G}\left[1 + \dfrac{4}{9\pi}(4-\alpha)\dfrac{c^3}{R_0^3}\right], \\[2mm] A = -\dfrac{N}{\pi G}. \end{array}\right\} \qquad \ldots\ldots(41)$$

In view of the orthogonality of P_2 and $P_0(\cos \vartheta) = 1$, the values of C and D do not enter into the result. The equation (39) becomes, with the use of (34), (36) and (41),

$$-2W = 2\pi R_0^2 \dfrac{N^2}{2G} \int_0^\pi \left[\dfrac{2(\alpha-1)}{4-\alpha}\left(1 + \dfrac{4c^3}{9\pi}\dfrac{4-\alpha}{R_0^3}\right)R_0 + \dfrac{2c^3}{9\pi}\dfrac{4-\alpha}{R_0^2}\right]\sin\vartheta d\vartheta,$$
$$\ldots\ldots(42)$$

$$W = -\dfrac{\pi N^2}{G}\left[2\dfrac{(\alpha-1)}{4-\alpha}R_0^3 + \dfrac{2\alpha}{3\pi}c^3\right]. \qquad \ldots\ldots(43)$$

In the absence of the crack the energy would have been

$$W_0 = -\dfrac{2\pi}{G}\dfrac{\alpha-1}{4-\alpha}N^2 R_0^3. \qquad \ldots\ldots(44)$$

The introduction of the crack therefore increases the free energy of the system by an amount

$$W_1 = W - W_0 = -\dfrac{2\alpha}{3G}N^2 c^3. \qquad \ldots\ldots(45)$$

Similarly, if an external stress $(M, -\tfrac{1}{2}M, -\tfrac{1}{2}M)$ is applied, the tractions S_{rr} and $S_{r\vartheta}$ for $r = R_0$ will be,

$$S_{rr} = M P_2(\cos\vartheta), \quad S_{r\vartheta} = -\tfrac{3}{2}M\sin\vartheta\cos\vartheta. \qquad \ldots\ldots(46)$$

The equality of the stresses (36) and (46), in conjunction with (38), leads to

$$W_1 = -\dfrac{2\alpha}{3G}M^2 c^3. \qquad \ldots\ldots(47)$$

A combination of these two types of stresses yields by the same method

$$W_1 = -\dfrac{2\alpha}{3G}(N+M)^2 c^3. \qquad \ldots\ldots(48)$$

Finally an additional stress of the type $(0, L, -L)$ will not affect the value of W_1.

§4. RESULT AND DISCUSSION

If a brittle body is acted upon by a uniform stress (P, Q, R), (48) shows that the presence of a circular crack of radius c in a plane normal to P alters the

free energy of the solid by an amount

$$W_1 = -\frac{2\alpha}{3G}P^2c^3 = -\frac{8(1-\sigma^2)}{3E}P^2c^3, \qquad \ldots\ldots(49)$$

irrespective of Q and R, provided P is tensile and c is appreciably smaller than the dimensions of the solid. On the other hand, the crack has a surface energy amounting to

$$W_2 = 2\pi c^2 T, \qquad \ldots\ldots(50)$$

where T represents the surface tension of the material. Thus the total free energy contribution due to the crack is

$$\Delta W = W_1 + W_2 = 2\pi c^2 T - \frac{8(1-\sigma^2)}{3E}P^2c^3. \qquad \ldots\ldots(51)$$

For varying c this expression has a maximum when

$$c = \frac{\pi E T}{2P^2(1-\sigma^2)}. \qquad \ldots\ldots(52)$$

If c is greater than this value, or alternatively if

$$P > P_c = \sqrt{\left(\frac{\pi E T}{2c(1-\sigma^2)}\right)}, \qquad \ldots\ldots(53)$$

the crack will become unstable and spread. If plane circular cracks of radius c are distributed at random in a brittle solid, rupture will be determined solely by the maximum tensile stress; the tensile strength in one direction will not be affected by the presence of smaller tensile (and, within limits, compressive) stresses at right angles to it.

Comparison of the result (53) with the plane case (1) and (2) shows that the tensile strengths differ by a factor ranging between $\frac{1}{2}\pi = 1\cdot57$ and $\pi/\sqrt{3} = 1\cdot81$. If the cracks are supposed to be plane ellipses, then for a given minor axis c the variation of the major axis from infinity to c affects the tensile strength of the material by that factor only. The tensile strength is thus primarily determined by the length of the minor axes of the cracks and must lie between the limits given by Griffith (1921 and 1924) and in equation (53).

REFERENCES

ANDRADE and MARTINDALE, 1936. *Phil. Trans.*, **235**, 69.
GRIFFITH, 1921. *Phil. Trans.*, **221**, 180.
GRIFFITH, 1924. *II Int. Congress Appl. Maths.*, Delft, p 55.
HARDING and SNEDDON, 1945. *Proc. Camb. Phil. Soc.*, **41**, 16.
HOBSON, 1931. *Theory of Spherical and Ellipsoidal Harmonics* (Cambridge : The University Press).
JOFFE and LEVITZKY, 1925. *Z. Phys.*, **31**, 567.
NEUBER, 1934. *Z. Ang. Math. Mech.*, **14**, 203.
REINKOBER, 1937. *Phys. Z.*, **37**, 112.
SNEDDON, 1946. *Proc. Roy. Soc.*, A (in the press).

THE THERMODYNAMIC PROPERTIES OF THE PRODUCTS OF HIGH-PRESSURE COMBUSTION

By J. CORNER,

Armament Research Department, Ministry of Supply

ABSTRACT. The internal energy, heat content, specific heats and equilibrium constants of the normal products of combustion are expanded as power series in the density. At pressures up to 40 tons/sq. in., corresponding to densities of about 0·35 gm./c.c., only the first three terms need be kept. Thus the pressure-dependence of the thermodynamic properties can be evaluated from a knowledge of the second and third virial coefficients of the various product gases. Tables are presented based on the most recent intermolecular forces, and covering the range from 1600° K. to 4000° K. These tables have found considerable application to internal ballistics, and give " covolumes " of propellants with a systematic error of less than 5%.

§1. INTRODUCTION

FROM the equation of state of a gas we can derive the contributions of gas-imperfection to thermodynamic properties such as its internal energy. In the application to the products of high-pressure combustion we desire to find the energy and heat content of the products of explosion, at temperatures from 1500 to 4000° K. and at densities up to 0·35 gm./c.c. Under these conditions it is sufficient to retain the second and third virial coefficients. Theoretical expressions for these are known in the case of the pure gases, and the arbitrary parameters can be chosen to fit the observed second virial coefficient at low temperatures. Hence we can find the equations of state of the pure gases at high temperatures, and so, by using thermodynamic formulae, we can find their internal energies and heat contents. Relating the virial coefficients of the mixture to those of the constituents, by means of a simple combination rule, we derive, finally, the properties of the product mixture and also the pressure-correction to the equilibrium constants which determine the composition.

The details of this programme are given in the present paper. A similar aim, with largely the same methods, has been pursued by Hirschfelder and his co-workers, in work which was given a restricted circulation while the present work was being carried out. The points of difference have been considered sufficiently important to warrant the completion of the present work. These differences lie chiefly in the treatment of carbon dioxide, and the extent to which one may regard the second virial coefficient as independent of temperature. There are also minor differences in the equations and parameters used.

The basis of the present work is the expression of the second virial coefficient of a gas as a simple function of the parameters of the intermolecular field. There is a general expression for the virial coefficient in the form of an integral over the intermolecular field at all distances; the difficulty has always been in the

discovery of an intermolecular potential of plausible form, with an integrable second virial coefficient. For molecules with spherical symmetry, we use a special case of the well-known Lennard-Jones potential (Lennard-Jones, 1924). This has been used in the discussion of nitrogen, carbon monoxide and hydrogen (§§ 4 and 5 of this paper). Steam is covered by Stockmayer's recent development (Stockmayer, 1941) of a potential which represents the field of a polar molecule with a short-range repulsion of spherical symmetry. We have also taken into account the work of Hirschfelder, McClure and Weeks (1942) and Margenau and Myers (1944) on steam. Carbon dioxide is an example of a cigar-shaped, non-polar molecule, whose virial coefficient has been studied by Corner (1946).

The most obvious application of our tables is to problems of internal ballistics, particularly to the testing of propellants in closed vessels. The standard thermochemical tables used in closed-vessel work in Britain are those of Pike (unpublished). The specific heats and equilibrium constants tabulated there refer to zero pressure, and the explosion temperatures, pressures and product compositions calculated from these tables are therefore incorrect. Of these quantities only the pressure can be measured experimentally; the theoretical pressure is adjusted by the introduction of a " covolume ", chosen to give the right pressure at normal loading densities. The theoretical temperatures and compositions are in error by unknown amounts.

The tables in the present paper are intended to be used in conjunction with tables such as those of Pike. They purport to give real pressures of explosion, and this is confirmed by the comparison with experimental results in § 12. To put it another way, we may say that these tables enable covolumes to be predicted. The tables also give true temperatures and compositions, both of which are difficult to test by experiment. At the same time this fact makes theoretical results all the more important in physical and chemical problems of internal ballistics, such as questions of erosion by heat transfer and chemical action.

§ 2. BASIS OF THE TABLES

We wish to find the internal energy and heat content of the products of combustion at pressures up to 40 tons/sq. in. and at temperatures of order $1500-4000° \text{K}$.* The values at zero density can be deduced from spectroscopic data. The pressure-corrections can be derived from the equations of state of the products, by thermodynamic formulae. This use of the equation of state to obtain theoretical specific heats is now common in chemical engineering, but we have been able to find only one published application at the temperatures encountered in internal ballistics. Burlot (1932) has calculated the specific heats of CO_2 and N_2 from an empirical equation of state, claimed to hold up to 100° c. and 300 atmospheres. This is far too slender a basis for even qualitative results under ballistic conditions. The only other early work which bears on this subject is that of Crow and Grimshaw (1931), who attempted to derive molecular radii from dielectric constants. When suitably corrected the results fitted the observed covolumes of non-flashless propellants. This has no more than an empirical

---

* The upper limit is settled by the explosion temperature of the hottest Service propellants, about 3700° K.; temperatures below 1500° K. cause difficulties with terms neglected in our formulae, and, in any case, the assumption of chemical equilibrium, necessary to the use of thermo-chemical tables, is violated in many processes of combustion ending at lower temperatures.

significance, for in fact the "covolume" may arise not only from the volume occupied by the molecules but also from the resulting change in the specific heats; the latter cannot be calculated from the rough equation of state used by Crow and Grimshaw. It is true that this change of the explosion temperature has been proved, from the present tables, to be negligible, but this is a result which could not have been deduced from an approximate equation of state.

The calculation of heat and energy contents requires an accurate equation of state. Experimental results do not go beyond 700° K., so there is a wide range to be covered by extrapolation before we reach the region 1500–4000° K., of interest in combustion. The extrapolation is, in fact, sufficiently wide to be unbridgeable in an empirical way. Fortunately there is now a considerable body of theory to help us, and it plays an essential part in our work. It is perhaps as well to emphasize that this theory has been verified and tested on the equations of state of gases at normal temperatures, so that, in effect, the theory is being used as an extrapolation formula with the backing of statistical mechanics. The work is straightforward, and the only doubtful point is whether our equations of state and specific heats are, in fact, sufficiently accurate. We shall describe various tests which have been applied. Meanwhile we may say that the basic data (equations of state of the pure gases up to 600–700° K.) and the theory are well established, and future modifications are likely to be small; also there seems to be little hope of working back from experiments at ballistic temperatures. Accordingly our results are not likely to be much altered by new experimental data at low temperatures and not likely to be superseded by a more direct method.

In a purely empirical extrapolation to high temperatures the only reason for confidence in the results would be the accuracy of fit to experiments at normal temperatures and possibly the successful prediction of covolumes. The essential basis of our theoretical extrapolation is the reasonable nature of the intermolecular forces assumed. For example, we treat N_2, CO and H_2 as having fields of spherical symmetry, for this assumption is not only sufficient to account for the low-temperature equation of state but is also reasonably consistent with what we know of their molecular structure. The same applies to H_2O, except that here we have used a field which is effectively that of a dipole with a short-range repulsion of spherical symmetry; a theory of such intermolecular fields was produced by Stockmayer in 1941. The other main product of propellant explosions is CO_2, which presented the only real difficulty in our work. This molecule has a cigar shape, and it is not plausible to use a field of spherical symmetry even though this gives a good fit to experiments at normal temperatures. We have therefore devised a molecular field which is, roughly speaking, of cigar or cylindrical shape, and we have found the equation of state of a gas of such molecules (Corner, 1946). This equation we have fitted to observations on CO_2 up to 300° c., to find the parameters of the potential for this molecule. From this point the calculation of heat and energy contents at high temperatures has followed the same course as for the simpler molecules.

It would be possible to apply the same methods to N_2, CO and H_2; the gain in accuracy is probably too small to justify at present the time that would be required.

The fact that the products of explosion are not perfect gases has an effect

on the equilibrium constants which determine the composition of the products. This effect, too, has been calculated from the equation of state by familiar thermodynamic formulae.

§3. EQUATION OF STATE OF A PURE GAS AND ITS THERMODYNAMIC PROPERTIES

Let P atmospheres be the pressure, V c.c. the volume, and $T°$ K. the temperature of a mole of pure gas. Up to the state of a saturated vapour the pressure can be expressed by the series

$$\frac{PV}{RT} = 1 + \frac{B}{V} + \frac{C}{V^2} + \frac{D}{V^3} + \ldots \ldots ,$$

where the coefficients B, C, D, etc. are functions only of T and the nature of the gas. R is $82 \cdot 06$ (Birge, 1941). B and C are the second and third virial coefficients respectively. We shall use

$$\frac{PV}{RT} = 1 + \frac{B}{V} + \frac{C}{V^2}, \qquad \ldots\ldots(1)$$

which is sufficiently accurate for our purposes. For example, at these temperatures B for nitrogen is of the order 30 c.c./mole, C of order 150 (c.c./mole)^2, so that at densities around $0 \cdot 35$ gm./c.c. the successive terms on the right of (1) are 1, $0 \cdot 4$ and $0 \cdot 02$. It is obvious that even at this density, which lies at the upper end of the range occurring in combustion problems, the neglect of higher terms in the expansion leads to errors of only $0 \cdot 1 \%$ in the pressure.

It may be worth pointing out that the three-term power-series in (1) is considerably more accurate * than the series

$$\frac{PV}{RT} = 1 + R(T)P + S(T)P^2,$$

which is often used in chemical engineering. The latter equation has the form $V = f(P, T)$, which is a virtue in that field. In ballistics the most convenient form is $P = f(V, T)$, which is fortunately a property of the more accurate equation (1).

Let $E(T, V)$ and $H(T, V)$ be the internal energy and the heat content of the mole of gas at temperature $T°$ K. and volume V c.c. In general,

$$\left(\frac{\partial E}{\partial V}\right)_T = T\left(\frac{\partial P}{\partial T}\right)_V - P,$$

and using (1) we find

$$E(T, V) - E(T, \infty) = -RT^2\left(\frac{B'}{V} + \frac{C'}{2V^2}\right), \qquad \ldots\ldots(2)$$

where dashes denote derivatives with respect to T. We write this as

$$E(T, V) = E_0(T) + \frac{E_1(T)}{V} + \frac{E_2(T)}{V^2}, \qquad \ldots\ldots(3)$$

where

$$E_1 = -RT^2 B'$$

and

$$E_2 = -\tfrac{1}{2}RT^2 C'. \qquad \Bigg\} \qquad \ldots\ldots(4)$$

---

* Cf. p. 203 of Beattie and Stockmayer (1940).

From $H = E + PV$ we have

$$H(T, V) = H_0(T) + \frac{H_1}{V} + \frac{H_2}{V^2}, \qquad \dots \dots (5)$$

where

$$H_1 = RT(B - TB')$$

and

$$H_2 = RT(C - \tfrac{1}{2}TC'). \qquad \Big\} \qquad \dots \dots (6)$$

This can be checked by the relation

$$\left(\frac{\partial H}{\partial P}\right)_T = V - T\left(\frac{\partial V}{\partial T}\right)_P.$$

Let A be the Helmholtz free energy of the system. Since $\left(\dfrac{\partial A}{\partial V}\right)_T = -P$ we find from (1),

$$A(T, V) = A(T) - RT \ln V + \frac{RTB}{V} + \frac{RTC}{2V^2}. \qquad \dots \dots (7)$$

Let C_V and C_P be the specific heats at constant volume and pressure respectively.

From $\left(\dfrac{\partial C_V}{\partial V}\right)_T = T\left(\dfrac{\partial^2 P}{\partial T^2}\right)_V$ it follows that

$$C_V(T, V) = C_{V, 0}(T) + \frac{C_{V, 1}}{V} + \frac{C_{V, 2}}{V^2}, \qquad \dots \dots (8)$$

where

$$C_{V, 1} = -RT(2B' + TB''), \qquad \dots \dots (9)$$

$$C_{V, 2} = -\tfrac{1}{2}RT(2C' + TC''). \qquad \dots \dots (10)$$

Also

$$C_p - C_V = T\left(\frac{\partial P}{\partial T}\right)_V \left(\frac{\partial V}{\partial T}\right)_P$$

$$= R\left\{1 + \frac{\alpha}{V} + \frac{\beta}{V^2}\right\}^2 \left\{1 + \frac{2B}{V} + \frac{3C}{V^2}\right\}^{-1}, \dots \dots (11)$$

where

$$\alpha = B + TB', \qquad \dots \dots (12)$$

$$\beta = C + TC'. \qquad \dots \dots (13)$$

We shall use these results when we have obtained theoretical formulae for B and C. To this we now proceed, dealing in turn with each of the major products of combustion.

§4. HYDROGEN

Consider a pair of similar molecules, whose intermolecular potential energy W depends only on r, the distance between their centres. Let the zero of W be taken at infinite separation of the molecules. It can be shown that in classical statistical mechanics

$$B = 2\pi N \int_0^\infty \left[1 - \exp\left\{-\frac{W}{kT}\right\}\right] r^2 dr, \qquad \dots \dots (14)$$

where N is Avogadro's number and k is Boltzmann's constant.

Lennard-Jones (1924) has integrated this as a convergent series when-

$$W(r) = \phi\left(\frac{mp}{m-p}\right)\left[\frac{1}{m}\left(\frac{r_0}{r}\right)^m - \frac{1}{p}\left(\frac{r_0}{r}\right)^p\right] , \quad \dots\dots(15)$$

and m and p are constants satisfying $m > p > 3$; r_0 and ϕ are parameters character-
istic of the molecule, being respectively the position and depth of the minimum
of the potential. The behaviour of many simple molecules is well represented
by the special case

$$W = \phi\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^6\right] \quad \dots\dots(16)$$

which we shall use for H_2, N_2 and CO. This potential is often referred to as the
Lennard-Jones (12, 6) potential.

In the literature there are three sets of (ϕ, r_0) for hydrogen. The earliest
is that of Lennard-Jones (1931), deduced from observed second virial coefficients;
Hirschfelder and Roseveare (1939) have derived ϕ and r_0 from a study of the
Beattie-Bridgeman equation. Both these analyses neglected quantum corrections,
which are appreciable for hydrogen right up to room temperature. Allowance
for these effects has been made by de Boer and Michels (1938). Table 1 compares
the second virial coefficient at high temperatures from these three sets of para-
meters. The column marked " Keyes " was computed from a semi-empirical
equation which was proposed by Keyes (1941) as a summary of the experimental
data. For extrapolation to these high temperatures it is less reliable than the
results of the full theoretical analysis, to which the semi-empirical equation
is an approximation. All four sets of B show the same trend with temperature
and good agreement in magnitude. We have used the most recent theoretical
values, those of de Boer and Michels, in the preparation of our tables. These
parameters are

$$\phi/k = 37°\cdot02; \ r_0 = 3\cdot28 \text{ A}. \quad \dots\dots(17)$$

Table 1. B (c.c./mole) of H_2 at high temperatures

Temp. (°K.)	L-J	H, R	B, M	Keyes
1500	16·2	13·5	16·4	15·7
2000	15·7	13·1	16·0	15·1
2500	15·2	12·7	15·6	14·7
3000	14·8	12·4	15·2	14·2
3500	14·4	12·1	14·8	13·8
4000	14·1	11·9	14·5	13·5

We have omitted all quantum corrections to the second virial coefficient.
These corrections are largest for light molecules at low temperatures, and in the
worst case within our range (hydrogen at 1500° K.) they rise to 1%. For the
other products the corrections are very small; for example, they amount to only
1 part in 2000 for nitrogen at 1500° K.

We come now to the third virial coefficient, C. Recently $C(T)$ has been

integrated for the intermolecular potential (16) by de Boer and Michels (1939) and Montroll and Mayer (1941). We define a new dimensionless variable by

$$\tau = kT/\phi. \qquad \ldots\ldots(18)$$

It can be shown that

$$C = \left\{ \frac{\sqrt{2}\pi N r_0}{3} \right\}^2 C^*(\tau) \qquad \ldots\ldots(19)$$

merely by dimensional analysis. Here C^* is a function of τ only. de Boer and Michels computed C^* for τ between 4 and 1·5 by numerical integration of a function which itself involved a numerical double integral; Montroll and Mayer started with the same formula for C^*, but transformed it into numerical integration of a function which has been computed by a single integration. We have plotted the two sets of results in figure 1. There is some disagreement between the two sets, especially at small values of $1/\tau$, which is just where we intend to use C^*. However, the Montroll-Mayer curve is the smoother, and
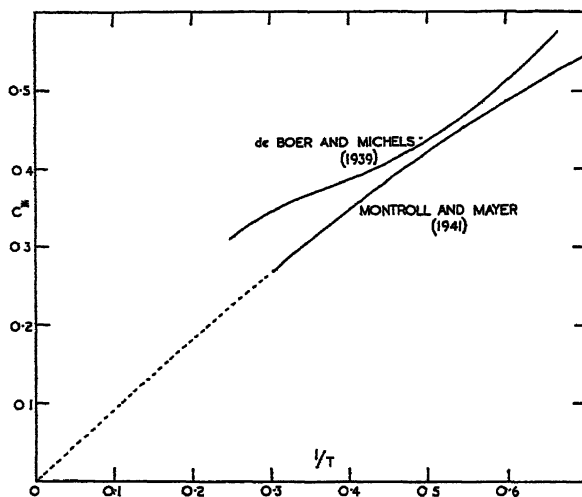
Figure 1. Alternative theoretical results for the coefficient C^*.

a comparison with the rather scattered experimental values of C^* is more favourable to Montroll and Mayer's curve. We shall assume their curve, and shall use it to extrapolate C^* towards $1/\tau = 0$, where $C^* = 0$. Various series of integral and fractional powers of $1/\tau$ have been used to represent the behaviour of C^*, but none of these differs significantly from

$$C^* = \frac{0\cdot9315}{\tau} - \frac{0\cdot0738}{\tau^2} + \frac{0\cdot212}{\tau^3} \qquad \ldots\ldots(20)$$

in the region of τ up to 40. This simple formula fits the curve of Montroll and Mayer to within the errors of reading from the latter, for τ greater than 1·5 (cf. figure 1). We have used (17), (18), (19) and (20) to get C and dC/dT for hydrogen at temperatures from 1500 to 4000° K. This corresponds to the range of τ from 40 to 100, where C is extremely small, of order 15 (c.c./mole)²; B is of order 15 c.c./mole.

Table 2 gives some idea of the size of the corrections caused by gas imperfections in hydrogen under typical ballistic conditions. These results were calculated from the tables given in §11 of the present paper, combined with those compiled by Dr. H. H. M. Pike for the ideal gas.

Table 2.   Gas imperfection in hydrogen at 3000° K. and 100 c.c./mole

	Ideal gas	From our tables	Percent difference
Pressure (atmos.)	2460	2839	15·4
$H-H$ (300° K., zero density), cal./mole	23500	24540	4·4
$E-E$ (300° K., zero density), cal./mole	17540	17670	0·8
C_V, cal./deg. mole	6·808	6·866	0·9
C_P, cal./deg. mole	8·795	8·813	0·2

§5. NITROGEN AND CARBON MONOXIDE

These two substances have molecules with many closely similar properties (cf. Grimm and Wolf, 1933), and in particular their equations of state for the gas phase are very much alike (Beattie and Bridgeman, 1928; Newitt, 1940). We shall treat both substances together, using the data on nitrogen, which are more extensive than those on carbon monoxide.

The most recent values of the parameters r_0 and ϕ for nitrogen are those of Corner (1940), based on Joule-Thomson data as well as second virial coefficients. The latter alone were used in the earlier work of Lennard-Jones (1931). Other values have been obtained by Hirschfelder and Roseveare (1939) from the constants of the Beattie-Bridgeman equation (CO and N_2) and from an analysis of Joule-Thomson data (N_2). Keyes (1941) has given a semi-empirical correlative equation. The resulting values of B are shown in table 3.   They are in good agreement,

Table 3. B (c.c./mole) of N_2 at high temperatures

Temp. (°K.)	Keyes	L-J	JC	H, R
1500	31·1	33·4	31·9	32·3
2000	31·5	34·3	32·6	33·1
2500	31·4	34·4	32·7	33·3
3000	31·0	34·3	32·6	33·2
3500	30·6	34·0	32·3	32·9
4000	30·2	33·7	32·0	32·6

and the most recent theoretical values lie near the mean of the others.   We have therefore, used the Corner parameters

$$\phi/k = 95° \cdot 3_3; \quad r_0 = 4 \cdot 10 \, \text{A}. \qquad \ldots\ldots(21)$$

The computation of B, C, E and H followed the same course as for hydrogen Table 4 shows the magnitude of the effects for nitrogen.

·Table 4.· Gas imperfection in nitrogen at 3000° K. and 100 c.c./mole

	Ideal gas	From our tables	Percent difference
Pressure (atmos.)	2460	3290	33·7
$H-H$ (300° K., zero density), cal./mole	22170	24290	9·6
$E-E$ (300° K., zero density), cal./mole	16800	16920	0·7
C_V, cal./deg. mole	6·876	7·018	2·1
C_P, cal./deg. mole	8·863	9·048	2·1

§6. STEAM .

This cannot be treated by the field of spherical symmetry assumed for nitrogen and hydrogen. Stockmayer (1941) has found the second virial coefficient for a potential which represents a dipole interaction superposed on the usual London attraction and short-range repulsion; both the latter terms were taken to be spherically symmetrical. He has applied the theory to the experimental data for steam, assuming the repulsive potential to have the form Ar^{-24}. This is a much higher index than the values, around 12, found for simple non-polar molecules. Hirschfelder, McClure and Weeks (1942) have repeated the comparison between experiment and the theory with the more plausible index of 12. We have compared the virial coefficients at 3200° K. from their parameters and those of Stockmayer; B is respectively 7·75 and 12·3 c.c./mole, TdB/dT is 9·1 and 11·6 c.c./mole, and $T^2 d^2B/dT^2$ is $-24·6$ and $-31·1$ c.c./mole. In computing our tables we have used the parameters found by Hirschfelder, McClure and Weeks.

Margenau and Myers (1944) have produced an intermolecular field for steam which they claim is more plausible than that of Stockmayer. We have therefore repeated all our calculations on steam, using the Margenau-Myers field, and have computed covolumes of propellants (cf. § 12) from both sets of tables. The Margenau-Myers field increases the error in the covolume from 4 to 6%. We have followed the indication of this empirical test, and in the tables of § 11 the values for steam are based on the work of Hirschfelder, McClure and Weeks.

There has been no theoretical investigation of the third virial coefficient for polar molecules. To estimate C we proceeded thus: the *theoretical* second virial coefficient from 1400 to 4000° K. was analysed in the usual way with the Lennard-Jones (12, 6) potential (16), and it was found that the virial coefficient could best be represented by the virial from (16) if the constants in the latter were given the values

$$\phi/k = 560°; \quad \frac{\sqrt{2}\pi N r_0^3}{3} = 26·1 \text{ c.c./mole.} \quad \ldots\ldots(22)$$

These are plausible values.* We have assumed that from 1400 to 4000° K. the third virial coefficient also is approximately the same as that from the potential (16) with the constants (22).

* For non-polar molecules the critical point is given by

$$V_c \simeq 1·48 \left(\frac{\sqrt{2}\pi N r_0^3}{3} \right) \text{ and } T_c \simeq 1·3\phi/k,$$

and in conjunction with (22) they would give $T_c = 750°$ K. and $V_c = 39$ c.c. mole. Observed values for water are 647° K. and 45 c.c./mole. This shows that the parameters (22) are of the right order of magnitude.

Table 5 shows the size of the corrections to thermodynamic data for steam.

Table 5.   Gas imperfection in steam at 3000° K. and 100 c.c./mole

	Ideal gas	From our tables	Percent difference
Pressure (atmos.)	2460	2664	8·3
$H-H$ (300° K., zero density), cal./mole	29930	29860	−0·2
$E-E$ (300° K., zero density), cal./mole	24570	24000	−2·3
C_V (cal./deg. mole)	11·107	11·246	1·3
C_P (cal./deg. mole)	13·094	13·566	3·6

§7. CARBON DIOXIDE

This is another molecule whose intermolecular field cannot be represented by terms with spherical symmetry.   Both long and short-range forces depend on the mutual orientation of the pair of molecules considered.   There does not appear to be any theoretical discussion of such a case in the literature.   The requirement is the second virial coefficient for molecules which are non-polar, roughly cylindrical, and stiff.

It is obvious that molecules which resemble cylinders form a very wide class, for not only the nuclear framework but also the nature of the nuclei in different parts of the molecule vary from one case to another.   The intermolecular field, therefore, has to satisfy conflicting requirements of generality and accuracy of representation, as well as the essential point that it must give an integrable expression for the second virial coefficient.   The latter requirement is not important in the case of spherical symmetry, since the integral has in that case only one variable of integration; numerical integration is always possible, though it may be laborious.   When the molecules have lower symmetry the second virial coefficient is given by an integral with at least four variables of integration, while the specification of the field itself requires at least one more parameter than in the case of spherical symmetry.   Numerical integration is therefore impracticable.

We shall explain in outline how we have tried to find an intermolecular potential which satisfies the three conditions of accuracy, generality and integrability. Details are being published elsewhere (Corner, 1946).   As a first step it was assumed that at any fixed relative orientation the intermolecular potential depended on the distance between the centres of the molecules in a way which has been found to give a good representation of the behaviour of molecules with spherical symmetry.   The second virial coefficient was thus reduced to a triple integral.   This potential has two parameters, R_0 and ϕ_0, which are respectively the position and depth of the minimum potential.   For cylindrical molecules ϕ_0 and R_0 are functions of the relative orientation.   In the next step it was necessary to find the dependence of $\phi_0{}^{\frac{1}{3}}$ and $R_0{}^3$ on the three angles which specify the orientation and on the length/breadth ratio of the molecule.   The first model investigated was a pair of centres of force separated by a certain distance; the potential assumed for the interaction of such centres of force was the usual potential for saturated spherical groups.   For various orientations of two such molecules

the mutual potential was evaluated numerically and the values of ϕ_0 and R_0 found by interpolation. This process was repeated with molecules represented by three and four centres of force distributed along the central axis. The three-centre model gave results differing considerably from the two-centre type, because the latter has a " waist " which cannot be expected as a feature of a real molecule. The four-centre type showed little difference in behaviour from the three-centre type. It was concluded that the four-centre model gave an adequate representation of how the normal intermolecular potential is altered by a change from spherical to cylindrical form. This four-centre model was then investigated for various " molecular lengths ". Formulae were found which fitted the variation of $\phi_0{}^{\ddagger}$ and $R_0{}^3$ with the three orientation variables and the length/diameter ratio of the model. Finally these formulae were substituted in the integral for the second virial coefficient and this was integrated by expansion in series.

It is thought that the potential energy so derived is likely to be a good approximation for most molecules of cylindrical type, while, of course, it may be considerably in error for special cases. The molecule is assumed to be of roughly the same nature at both ends, and there are certainly many molecules which do not satisfy this condition. It is, however, true for carbon dioxide, and the other conditions of a fairly stiff nuclear framework and zero resultant dipole moment are also satisfied in this case.

In this way we have computed tables which give the second virial coefficient in terms of three variables: r_0 and ϕ are related to the depth and position of the minimum potential at a standard orientation; the third variable is effectively the length of the nuclear framework of the molecule. As this length tends to zero, r_0 and ϕ become the parameters already defined for spherically symmetrical systems in §4. We have also computed tables for the Joule-Thomson coefficient at low pressure, μ_0; the connection between this and the second virial coefficient was pointed out by Whitelaw (1934), and has been applied to the determination of intermolecular forces by Hirschfelder, Ewell and Roebuck (1938).

In applying these tables to carbon dioxide we took the length of the nuclear framework to be 2·32 A. (Adel and Denison, 1935). The most recent virial data are those of Michels and Michels (1935), from 0 to 150° c. They agree to within 4 c.c./mole with the virial coefficients deduced from Amagat's data (1893) by Bridgeman (1927). The Joule-Thomson coefficient has been found by Roebuck, Murrell and Miller (1942) over a wide range of pressure and from −50 to 300° c. We have extrapolated to zero pressure the values at temperatures over 0° c. ; below this temperature the extrapolation is rather uncertain.

The specific heat at zero density, $C_P{}^\circ$, which was needed in the analysis, was based on Kassel's work (Kassel, 1934). The best values of ϕ and r_0 were found to be

$$\phi/k = 198°; \quad \frac{\sqrt{2}\pi N r_0{}^3}{3} = 110 \text{ c.c.} \qquad \ldots\ldots(23)$$

This potential gives second virial coefficients with errors of order 2 c.c./mole, i.e. about 2% at room temperatures; the error in $\mu_0 C_P{}^\circ$ rises to about 10 c.c./mole at low temperatures, around 0°c., but is of order 5 c.c./mole (2%) over most of the temperature-range. Quantum corrections were included in this comparison.

At high temperatures the second virial coefficient becomes the same as that of the spherically symmetrical potential (16) with the same values of ϕ and r_0; this was arranged by the definitions of these parameters. For carbon dioxide the difference between the two virial coefficients is less than 1% at 1400° K. In preparing our tables we have therefore neglected the correction to the virial coefficient of the spherical potential. Hence we may compare the set (23) with others which have been derived from fields with spherical symmetry. Hirschfelder and Roseveare (1939) found (175°, 131 c.c.) from the Beattie-Bridgeman constants for carbon dioxide. Hirschfelder, McClure, Curtiss and Osborne (unpublished) have derived (185°, 120 c.c.) from isotherms of Michels, Blaisse and Michels (1937) (at medium rather than low pressures) and have checked these values against the Joule-Thomson data of Roebuck, Murrell and Miller (1942).

This spherically-symmetrical potential gives as good a fit to the low-temperature gas data as our potential with cylindrical symmetry, but we think the latter gives a more reliable extrapolation to high temperatures. Table 6 shows some high-temperature virial coefficients for these potentials.

Table 6.   Second virial coefficients of CO_2 at 3000° K.

Potential	B (c.c./mole)
(23)	56·2
Hirschfelder and Roseveare	60·3
(185°, 120 c.c.)	62·1

Since, at high temperatures, the theoretical second virial coefficient approximates to that of the Lennard-Jones (12, 6) potential, it is reasonable to assume that the third virial coefficient also can be represented by the results of Montroll and Mayer, calculated from a (12, 6) potential. We have therefore used (18), (19) and (20), with the constants (23), in finding the third virial.

Table 7.   Gas imperfection for CO_2 at 3000° K. and 100 c.c./mole

	Ideal gas	From our tables	Percent difference
Pressure (atmos.)	2460	4025	63·6
$H-H$ (300° K., zero density), cal./mole	36710	40240	9·6
$E-E$ (300° K., zero density), cal./mole	31350	31080	−0·9
C_V (cal./deg. mole)	13·020	13·277	2·0
C_P (cal./deg. mole)	15·007	15·567	3·7

§8. COMBINATION RULES

The discussion of §3 referred to a pure gas. For a mixture at low densities it is equally permissible to write

$$\frac{PV}{RT} = 1 + \frac{B_1}{V} + \frac{C_1}{V^2}, \qquad \ldots\ldots(24)$$

where V is the volume occupied by the gas of N molecules, N being Avogadro's number; B_1 and C_1 are functions of temperature and composition. The internal

energy and other thermodynamic data are still given by equations similar to those in §3, with the differentiations carried out at constant composition. The zero of internal energy, for example, refers to *the same* gases at zero density and 300° K.; C_P is the specific heat at constant pressure, *keeping the composition fixed.*

Coming now to the mixture of product gases, we consider the products of 1 gm., namely $[CO_2]$ moles of CO_2, $[CO]$ of CO, etc. Let

$$[CO_2] + [CO] + \ldots\ldots = n, \qquad \ldots\ldots(25)$$

and let V c.c. be the volume occupied by these gases. Then the molar volume of the mixture is (V/n) c.c., which is the volume to be used in equation (24). Hence

$$\frac{PV}{nRT} = 1 + \frac{nB_1}{V} + \left(\frac{n}{V}\right)^2 C_1. \qquad \ldots\ldots(26)$$

We must now obtain B_1 from the second virial coefficients of the pure gases. Several such combination rules are in use (cf. a review by Beattie and Stockmayer, 1940), but for our purposes only the simplest can be considered. This is the linear sum rule:

$$nB_1 = [CO_2]B(CO_2) + [CO]B(CO) + \ldots \qquad \ldots\ldots(27)$$

We use also the analogous formula:

$$nC_1 = [CO_2]C(CO_2) + [CO]C(CO) + \ldots \qquad \ldots\ldots(28)$$

From these and the formulae of §3, we arrive at

$$E(T, V) = \Sigma[CO]\left\{E_{CO,0}(T) + \frac{n}{V}E_{CO,1}(T) + \left(\frac{n}{V}\right)^2 E_{CO,2}(T)\right\} \qquad \ldots\ldots(29)$$

for the internal energy relative to the same gases at 300° K. and zero density. $E_{CO,0}(T)$ is the internal energy of a mole of CO at zero density and temperature T; $E_{CO,1}$ and $E_{CO,2}$ are the corrections from the second and third virial coefficients. The summation is over all constituents of the mixture. The explicit formulae for $E_{CO,1}$ and $E_{CO,2}$, given in equation (4), are

$$E_{CO,1} = -RT^2B'(CO), \qquad \ldots\ldots(30)$$

$$E_{CO,2} = -\tfrac{1}{2}RT^2C'(CO). \qquad \ldots\ldots(31)$$

Similarly,

$$H(T, V) = \Sigma[CO]\left\{H_{CO,0}(T) + \left(\frac{n}{V}\right)H_{CO,1}(T) + \left(\frac{n}{V}\right)^2 H_{CO,2}(T)\right\} \qquad \ldots\ldots(32)$$

with

$$H_{CO,1} = RT\{B(CO) - TB'(CO)\}, \qquad \ldots\ldots(33)$$

$$H_{CO,2} = \tfrac{1}{2}RT\{2C(CO) - TC'(CO)\}. \qquad \ldots\ldots(34)$$

Formulae for C_V and C_P can be derived from those in §3.

In §11 the quantities E_1, E_2, H_1 and H_2 are tabulated for the main products of combustion at high pressures.

The other two well-known combination rules are the "square-root" and Lorentz rules (cf. Beattie and Stockmayer, 1940). These involve squares and products of the mole numbers such as $[CO_2]$, and greatly increase the labour

in the evaluation of pressures, internal energies, and heat contents. It is doubtful, too, whether the real accuracy of our calculations for the individual gases is sufficient to justify a superstructure of elaborate combination rules. To examine the order of magnitude of the changes involved, we have calculated the second virial coefficient for the products of a typical cordite. The square-root rule gave a value smaller by $8 \cdot 5\%$, the Lorentz rule a value smaller by $5 \cdot 5\%$. Probably the latter is the more accurate. The latter change corresponds to a 2% change in pressure at 30 tons/sq. in. The sign of the change indicates that the agreement with closed-vessel experiments would be made worse by the use of the more complicated rules.

The third virial coefficient has such a small effect on our thermochemical calculations that there is no need to look for a combination rule more accurate than the linear sum. This is also quite plausible.

§ 9. THE EQUILIBRIUM CONSTANT FOR THE WATER-GAS REACTION

From (27), (28) and (7) the Helmholtz free energy of 1 gm. of the mixture is

$$A = \Sigma[CO]A_{co}(T) + \Sigma[CO]RT \ln\{[CO]/V\} + \frac{nRT}{V}\Sigma[CO]B(CO)$$

$$+ \left(\frac{n}{V}\right)^2 \frac{RT}{2}\Sigma[CO]C(CO). \qquad \ldots\ldots(35)$$

The chemical potential of CO in the mixture is

$$\mu_{co} = \left\{\frac{\partial A}{\partial[CO]}\right\}_{T, V, [CO_s]}, \ldots\ldots$$

$$= A_{co}(T) + RT \ln\{[CO]/V\} + RT + \frac{nRT}{V}B(CO) + \frac{RT}{V}\Sigma[CO]B[CO]$$

$$+ \left(\frac{n}{V}\right)^2 \frac{RT}{2} C(CO) + \frac{nRT}{V^2}\Sigma[CO]C(CO). \qquad \ldots\ldots(36)$$

The fugacity $P^*(CO)$ is therefore given by

$$RT \ln P^*(CO) = RT \ln\{[CO]RT/V\} + \frac{nRT}{V}B(CO) + \frac{RT}{V}\Sigma[CO]B(CO)$$

$$+ \left(\frac{n}{V}\right)^2 \frac{RT}{2} C(CO) + \frac{nRT}{V^2}\Sigma[CO]C(CO),$$

and hence

$$P^*(CO) = \frac{RT[CO]}{V}\exp\left[\frac{nB(CO)}{V} + \left(\frac{n}{V}\right)^2\frac{C(CO)}{2} + \frac{1}{V}\Sigma[CO]B(CO)\right.$$

$$\left. + \frac{n}{V^2}\Sigma[CO]C(CO)\right]. \qquad \ldots\ldots(37)$$

The equilibrium of the water-gas reaction is determined by

$$\frac{P^*(CO)P^*(H_2O)}{P^*(CO_2)P^*(H_2)} = K_0(T), \qquad \ldots\ldots(38)$$

where K_0 is the equilibrium constant at low densities.

Hence

$$\frac{[CO][H_2O]}{[CO_2][H_2]} = K_0(T)\exp\left\{-\frac{n\Delta B}{V} - \frac{n^2\Delta C}{2V^2}\right\}, \qquad \ldots\ldots(39)$$

where

$$\Delta B = B(CO) + B(H_2O) - B(CO_2) - B(H_2) \qquad \ldots\ldots(40)$$

and

$$\Delta C = C(CO) + C(H_2O) - C(CO_2) - C(H_2). \qquad \ldots\ldots(41)$$

These are negative and of the same order of magnitude as the virial coefficients of carbon monoxide. The deviations from the perfect gas law therefore increase the effective equilibrium constant by as much as 50% at the highest densities attained in ballistics. ΔB and ΔC are tabulated in § 11.

§ 10. MINOR PRODUCTS

The ratios of the major products are determined by the equilibrium constant of the water-gas reaction and the composition of the fuel. The minor products [OH], [H], and [NO] are determined by the following equations

$$\frac{[OH]\sqrt{[H_2]}}{[H_2O]} = \sqrt{\frac{V}{RT}}\,K_1(T)\exp\left[\frac{n}{V}\{B(H_2O) - B(OH) - \tfrac{1}{2}B(H_2)\}\right], \ldots\ldots(42)$$

$$\frac{[NO][H_2]}{[H_2O]\sqrt{[N_2]}} = \sqrt{\frac{V}{RT}}\,K_2(T)\exp\left[\frac{n}{V}\{B(H_2O) + \tfrac{1}{2}B(N_2) - B(NO) - B(H_2)\}\right],$$
$$\ldots\ldots(43)$$

$$[H]/\sqrt{[H_2]} = \sqrt{\frac{V}{RT}}\,K_3(T)\exp\left[\frac{n}{V}\{\tfrac{1}{2}B(H_2) - B(H)\}\right]. \qquad \ldots\ldots(44)$$

K_1, K_2 and K_3 are functions listed in the literature.

The terms involving third virial coefficients have been neglected. We need B for OH, H and NO. We must keep in mind the fact that the pressure-corrections in (42), (43), and (44) will have only a small influence on thermochemical calculations, and therefore some attention may be paid to convenience in computation. As reasonable rough values for the second virial coefficients at high temperatures we take $B(H) \simeq 10$ c.c./mole ; $B(OH) \simeq 20$ c.c./mole; $B(NO) \simeq 30$ c.c./mole and then

$$B(H_2O) - B(OH) - \tfrac{1}{2}B(H_2) \simeq -20 \text{ c.c./mole},$$

$$B(H_2O) + \tfrac{1}{2}B(N_2) - B(NO) - B(H_2) \simeq -20 \text{ c.c./mole},$$

while $\tfrac{1}{2}B(H_2) - B(H)$ is only a few c.c./mole over the temperature ranges at which these species are present. Hence we can write, with sufficient accuracy,

$$\frac{[OH]\sqrt{[H_2]}}{[H_2O]} = \sqrt{\frac{V}{RT}}\,K_1(T)\exp(-20\,n/V), \qquad \ldots\ldots(45)$$

$$\frac{[NO][H_2]}{[H_2O]\sqrt{[N_2]}} = \sqrt{\frac{V}{RT}}\,K_2(T)\exp(-20\,n/V), \qquad \ldots\ldots(46)$$

$$[H]/\sqrt{[H_2]} = \sqrt{\frac{V}{RT}}\,K_3(T). \qquad \ldots\ldots(47)$$

It is convenient that two of the pressure-corrections are the same and the third is unity.

Of the other products of propellant explosions, O_2, O, and N occur in such small quantities that the relevant equilibrium constants need not be corrected for gas-imperfection. Methane and ammonia also occur in only small amounts. There would be no difficulty in calculating high-temperature virial coefficients for these gases; methane is effectively spherical and ammonia is covered by Stockmayer's work (Stockmayer, 1941).

In practical calculations based on our tables we have neglected the contribution of the minor products to the virial coefficients of the mixture. Taking the rough values used above, we have found that the inclusion of these terms would change the pressure by only 0·2% in the worst case. Similarly the minor products make only a small contribution to the pressure-dependent parts of the internal energy and heat content of the product mixture.

§ 11. TABLES OF RESULTS

The notation has been explained in §§ 3 and 9. The number of places kept in these tables is greater than the real accuracy, and was chosen to give products of combustion smooth to $\frac{1}{2} \times 10^{-5}$ mole/gm. and temperatures smooth to one or two degrees.

(a) Tables for pressures

Temp. (°K)	Table of B (c.c./mole)				Table of C (c.c./mole)^2			
	H_2	N_2, CO	CO_2	H_2O	H_2	N_2, CO	CO_2	H_2O
1600	16·4	32·1	45·7	−4·2	20	210	1385	220
1700	16·3	32·3	47·3	−2·5	20	200	1305	210
1800	16·2	32·4	48·7	−1·1	20	190	1235	195
1900	16·1	32·6	49·9	+0·2	20	180	1170	185
2000	16·0	32·6	50·9	1·2	15	170	1110	175
2100	15·9	32·7	51·8	2·2	15	160	1055	170
2200	15·8	32·7	52·6	3·0	15	155	1010	160
2300	15·7	32·8	53·2	3·7	15	150	965	155
2400	15·6	32·8	53·8	4·4	15	140	925	145
2500	15·6	32·8	54·4	5·0	15	135	885	140
2600	15·5	32·7	54·8	5·5	15	130	855	135
2700	15·4	32·7	55·3	6·0	15	125	825	130
2800	15·3	32·7	55·6	6·4	10	120	795	125
2900	15·3	32·6	56·0	6·8	10	120	765	120
3000	15·2	32·6	56·2	7·1	10	115	740	120
3100	15·1	32·6	56·5	7·5	10	110	720	115
3200	15·0	32·5	56·7	7·7	10	105	695	110
3300	15·0	32·4	56·9	8·0	10	105	675	105
3400	14·9	32·4	57·1	8·3	10	100	650	105
3500	14·8	32·3	57·3	8·5	10	95	635	100
3600	14·8	32·3	57·4	8·7	10	95	615	100
3700	14·7	32·2	57·5	8·9	10	90	600	95
3800	14·7	32·2	57·6	9·1	10	90	585	95
3900	14·6	32·1	57·7	9·3	10	85	570	90
4000	14·5	32·0	57·8	9·4	10	85	555	90

Reference to the tables of E_1 will show that the values for the different gases include both signs and accordingly the summation over all constituents tends to be small. E itself is of order 600–1000 cal./gm., and of this about -3 cal./gm. comes from the E_1 terms and about $+1$ cal./gm. from the E_2 terms (both these figures refer to pressures of order 20 tons/sq. in.). In the calculations of internal ballistics H is of order 600–1000 cal./gm., while H_1 terms contribute

(b) Tables for equilibrium constants
Definitions in equations (39)–(41)

Temperature ($^\circ$K.)	$-\Delta B$ (c.c./mole)	$-\frac{1}{2}\Delta C$ (c.c./mole)^2
1600	34·2	490
1700	33·8	460
1800	33·5	435
1900	33·2	410
2000	33·0	390
2100	32·8	370
2200	32·6	355
2300	32·5	340
2400	· 32·4	325
2500	32·2	310
2600	32·1	300
2700	32·0	290
2800	31·9	280
2900	31·8	270
3000	31·7	260
3100	31·6	255
3200	31·5	245
3300	31·4	235
3400	31·4	230
3500	31·3	225
3600	31·2	215
3700	31·1	210
3800	31·1	205
3900	31·0	200
4000	30·9	195

about 40 cal./gm. and the H_2 terms 4 cal./gm. at 20 tons/sq. in. This demonstrates how small are the pressure-corrections to the internal energy, compared with those to the heat content. As a corollary, calculations without pressure-corrections give temperatures of uncooled explosion correct to within a few degrees; on the other hand, the temperature of burning at constant pressure is really lower than that computed in such a simple manner, to the extent of 150° c. at 20 tons/sq. in.

C_{V_2} has maximum values of less than 40 (cal./deg.)(c.c./mole)^2 for H_2O, 20 for CO_2, 1 for N_2 and CO, and 0·1 for H_2. It has therefore not been tabulated. β, defined in (13), is less than 6 (c.c./mole)^2 for H_2O, 4 for CO_2, 2×10^{-4} for N_2 and CO, and 5×10^{-5} for H_2.

(c) Tables for internal energies and heats
Definitions in equations (3)–(6)

Temp. (°K.)	$E_1/100$ (cal.–c.c./mole)				$H_1/100$ (cal.–c.c./mole)			
	H_2	N_2, CO	CO_2	H_2O	H_2	N_2, CO	CO_2	H_2O
1600	45	−110	−870	−935	565	910	585	−1065
1700	50	− 95	−840	−895	600	995	760	− 975
1800	55	− 80	−810	−855	635	1080	930	− 895
1900	65	− 70	−785	−825	670	1160	1100	− 820
2000	70	− 55	−755	−795	705	1240	1265	− 745
2100	75	− 40	−730	−770	740	1325	1430	− 680
2200	80	− 25	−700	−745	775	1405	1595	− 615
2300	90	− 15	−675	−725	805	1485	1760	− 555
2400	95	0	−645	−705	840	1560	1920	− 495
2500	100	15	−620	−685	875	1640	2080	− 440
2600	105	25	−595	−665	905	1720	2240	− 385
2700	110	40	−565	−650	940	1795	2400	− 330
2800	120	55	−540	−635	970	1870	2555	− 280
2900	125	65	−515	−620	1005	1950	2710	− 230
3000	130	80	−490	−605	1035	2025	2865	− 180
3100	135	95	−465	−590	1065	2100	3015	− 135
3200	140	105	−440	−580	1100	2170	3170	− 85
3300	145	120	−415	−565	1130	2245	3320	− 40
3400	150	130	−390	−555	1160	2320	3470	+ 5
3500	160	145	−365	−540	1190	2395	3620	50
3600	165	155	−340	−530	1220	2465	3770	90
3700	170	170	−315	−520	1250	2540	3920	135
3800	175	185	−290	−510	1280	2610	4065	175
3900	180	195	−265	−500	1310	2680	4210	220
4000	185	210	−240	−490	1340	2750	4355	260

At all temperatures in this range use

	H_2	N_2, CO	CO_2	H_2O
$10^{-4}E_2$ (cal. (c.c./mole)^2)	3	34	220	35
$10^{-4}H_2$ (cal. (c.c./mole)^2)	10	101	661	106

(*d*) Tables for specific heats

Symbols defined in equations (8) – (13).

Temp. (°K.)	C_{V_1} (cal.–c.c./mole–deg.)				α (c.c./mole)			
	H₂	N₂, CO	CO₂	H₂O	H₂	N₂, CO	CO₂	H₂O
1600	6·4	14·3	29·4	43·8	15·0	35·6	73·1	25·2
1700	6·3	14·1	28·9	38·2	14·8	35·1	72·2	23·9
1800	6·3	14·0	28·4	33·9	14·6	34·7	71·4	22·9
1900	6·2	13·9	28·0	30·3	14·4	34·4	70·6	22·0
2000	6·2	13·8	27·7	27·4	14·2	34·0	69·9	21·3
2100	6·1	13·7	27·4	24·9	14·1	33·7	69·2	20·6
2200	6·1	13·7	27·1	22·9	14·0	33·4	68·6	20·1
2300	6·1	13·6	26·9	21·1	13·8	33·0	68·0	19·6
2400	6·0	13·5	26·7	19·6	13·7	32·7	67·4	19·1
2500	6·0	13·4	26·5	18·3	13·6	32·5	66·9	18·7
2600	5·9	13·4	26·3	17·2	13·5	32·2	66·3	18·4
2700	5·9	13·3	26·2	16·2	13·3	32·0	65·8	18·1
2800	5·9	13·2	26·0	15·4	13·2	31·7	65·4	17·8
2900	5·8	13·1	25·8	14·6	13·1	31·5	64·9	17·5
3000	5·8	13·1	25·7	13·9	13·0	31·3	64·5	17·3
3100	5·8	13·0	25·6	13·3	12·9	31·0	64·0	17·1
3200	5·7	13·0	25·5	12·8	12·8	30·8	63·6	16·9
3300	5·7	12·9	25·4	12·3	12·7	30·6	63·2	16·7
3400	5·7	12·9	25·3	11·8	12·7	30·4	62·9	16·5
3500	5·6	12·8	25·1	11·4	12·6	30·3	62·5	16·3
3600	5·6	12·8	25·0	11·0	12·5	30·1	62·1	16·1
3700	5·6	12·7	24·9	10·7	12·4	29·9	61·8	16·0
3800	5·6	12·7	24·9	10·4	12·4	29·7	61·5	15·9
3900	5·5	12·6	24·8	10·1	12·3	29·6	61·1	15·7
4000	5·5	12·6	24·7	9·9	12·2	29·4	60·8	15·6

§12. COMPARISON WITH EXPERIMENT

We have made a detailed study of the products of explosion of cordite " X ".

(a) *Explosion of " X " without performance of external work*

The following numerical data were taken as *exact*: atomic composition {C} = 2238; {H} = 3015; {N} = 1044; {O} = 3468, all in 10^{-5} gm.-atom/gm. ; heat of formation of propellant at constant volume = 470 cal./gm. From these data we have calculated the products of explosion, first from Dr. Pike's tables alone, and then with pressure-corrections from the tables given earlier in this paper. From the pressure of the products, we derive the " covolume " of the propellant by the definition

$$P(V-\eta)=nRT_0, \qquad \dots\dots(48)$$

where T_0 and n have been calculated at 4000 atmospheres from tables without pressure-corrections.

Table 8 shows the theoretical temperature, products, pressure and covolume at various product densities. In figures 2 and 3 are shown the behaviour of T_0, and the variation of one of the main products. T_0 differs by no more than 3° c. from that calculated without pressure-corrections. The main products are considerably different. The most interesting result in table 8 is the " theoretical

covolume " η. At pressures over 5 tons/sq. in., η decreases with increase of density. The accepted value for this cordite is 0·94 c.c./gm., determined by closed-vessel firings at 27 tons/sq. in. Our value at this pressure is 5% too small.
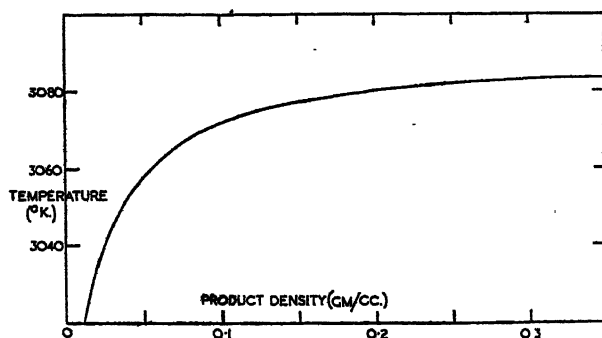
Figure 2. Product temperature in uncooled explosion of " X " without performance of external work.
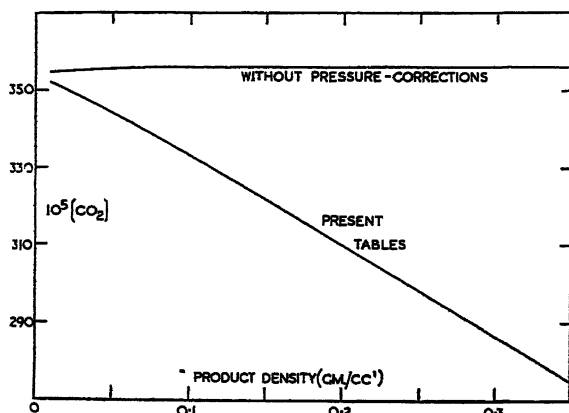
Figure 3. Variation of a major product with density in explosion of " X " without performance of external work.

As the maximum pressure is reduced below 5 tons/sq. in. the covolume begins to decrease again, until it becomes negative at pressures of order 1 ton/sq. in. This behaviour is due to the increasing dissociation causing low temperatures of explosion.

From our computations for the products of " X " at temperatures from 2000 to 3100° K., it can be shown that

$$P = \frac{nRT}{V-b} \qquad \dots\dots(49)$$

is, at each temperature, a less accurate representation than

$$P = \frac{nRT}{V}\left(1 + \frac{\beta}{V}\right). \qquad \dots\dots(50)$$

This had been predicted before the present tables were started, as it follows immediately from an inspection of the ratio (third virial coefficient)/(second virial coefficient)^2 at high temperatures. Using the best values of b and β, the

Table 8. Explosion of " X " at constant volume

Product density (gm./c.c.)	0·01	0·05	0·1	0·2	0·3	0·35
T_0, (°K.)	3018	3058	3073	3081	3081	3083
[CO$_2$] in 10^{-5} mole/gm.	352	343	333	310	286	275
[CO]	1886	1895	1905	1928	1952	1963
[H$_2$O]	853	873	888	914	938	951
[H$_2$]	628½	622	610	587	563	552
[N$_2$]	521½	521½	522	522	522	522
[OH]	23	12	8½	5	5	4
[H]	27	14	10	7	6	5
[NO]	1	1	½	½	0	0
$10^5 n$	4292	4282	4277	4274	4272	4272
P (atmos.)	107·5	568·2	1204	2672	4407	5381
P (tons/sq. in.)	0·705	3·73	7·90	17·53	28·91	35·30
Covolume, c.c./gm.	−0·5	0·98	1·02	0·96	0·88	0·85

errors in (49) and (50) rise to the order of 2% and 0·7% respectively. Both b and β vary little with T, certainly by less than 1% over the range 2000–3000°K.

Our tables have been applied with success to the calculation of the behaviour of other propellants, and show a systematic error of about 4% in the covolume.

(b) *Explosion of " X " under constant pressure*

The results of computations from our tables are shown in table 9 and figures 4 and 5. The temperature of the products, T_m, has a maximum just below 1 ton/sq. in., decreasing at smaller pressures because of the increasing dissociation
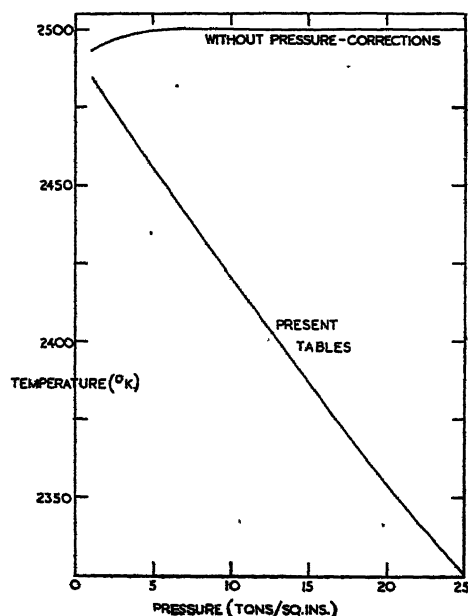
Figure 4. Theoretical product temperatures for explosion of " X " under constant-pressure conditions.
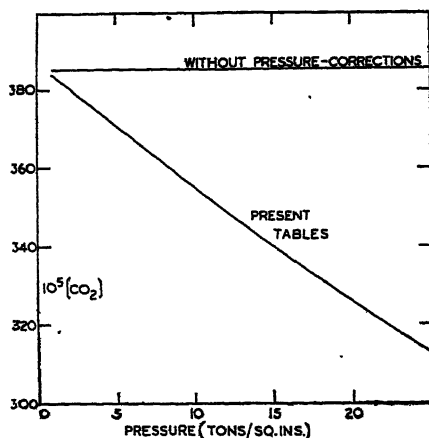
Figure 5. Composition of products of " X " under constant-pressure conditions.

and decreasing at higher pressures because of the increasing heat content of the imperfect gases. At 25 tons/sq. in. our T_m is about 180° c. lower than that calculated without pressure-corrections.

Table 9. Explosion of " X " under constant pressure

Pressure (atmos.)	88·2	457·8	956·6	2074·7	3351	4048
,, (tons/sq. in.)	0·58	3·00	6·28	13·61	21·99	26·56
Density of products (gm./c.c.)	0·01	0·05	0·1	0·2	0·3	0·35
T_m (°K.)	2488	2472	2448	2396	2343	2317
$[CO_2]$ in 10^{-5} mole/gm.	385	377	366	344	321	309
$[CO]$	1853	1861	1872	1894	1917	1929
$[H_2O]$	843	852	864	886	909	921
$[H_2]$	662	654	643	621	598	587
$[N_2]$	522	522	522	522	522	522
$[OH]$	2	1	0	0	0	0
$[H]$	4	2	1	1	1	0
$10^5 n$	4271	4269	4268	4268	4268	4268

ACKNOWLEDGMENTS

REFERENCES

ADEL and DENISON, 1935. *Phys. Rev.*, **48**, 516.
AMAGAT, 1893. *Ann. chim. phys.*, **29**, 109.
BEATTIE and BRIDGEMAN, 1928. *J. Amer. Chem. Soc.*, **50**, 3151.
BEATTIE and STOCKMAYER, 1940. *Rep. Progr. Phys.*, **7**, 195.
BIRGE, 1941. *Rev. Mod. Phys.*, **13**, 233.
DE BOER and MICHELS, 1938. *Physica*, **5**, 945 ; 1939. *Ibid.*, **6**, 97.
BRIDGEMAN, 1927. *J. Amer. Chem. Soc.*, **49**, 1130.
BURLOT, 1932. *Mem. Poudres*, **25**, 314.
CORNER, 1940. *Trans. Faraday Soc.*, **36**, 780 ; 1946. In the press.
CROW and GRIMSHAW, 1931. *Phil. Trans.*, A, **230**, 39.
GRIMM and WOLF, 1933. *Handb. Physik*, **24** (2nd edition), pt. 1, 977.
HIRSCHFELDER, EWELL and ROEBUCK, 1938. *J. Chem. Phys.*, **6**, 205.
HIRSCHFELDER, McCLURE and WEEKS, 1942. *J. Chem. Phys.*, **10**, 201.
HIRSCHFELDER and ROSEVEARE, 1939. *J. Phys. Chem.*, **43**, 15.
KASSEL, 1934. *J. Amer. Chem. Soc.*, **56**, 1838.
KEYES, 1941. *Temperature* (Amer. Inst. Phys., N.Y.), p. 45.
LENNARD-JONES, 1924. *Proc. Roy. Soc.*, A, **106**, 463 ; 1931. *Proc. Phys. Soc.*, **43**, 461.
MARGENAU and MYERS, 1944. *Phys. Rev.*, **66**, 307.
MICHELS, BLAISSE and MICHELS, 1937. *Proc. Roy. Soc.*, A, **160**, 358.
MICHELS and MICHELS, 1935. *Proc. Roy. Soc.*, A, **153**, 201.
MONTROLL and MAYER, 1941. *J. Chem. Phys.*, **9**, 626.
NEWITT, 1940. *High-pressure plant and fluids at high pressures*, p. 191.
ROEBUCK, MURRELL and MILLER, 1942. *J. Amer. Chem. Soc.*, **64**, 400.
STOCKMAYER, 1941. *J. Chem. Phys.*, **9**, 398.
WHITELAW, 1934. *Physica*, **1**, 749.

ON THE ZERNIKE PHASE-CONTRAST TEST

By E. H. LINFOOT,

University of Bristol

ABSTRACT. A diffraction theory of the Zernike test is developed for the case where the dimensions of the light source are negligibly small, and applied to estimate the sensitiveness of the test for detecting local zones in an otherwise true mirror. The effects of varying the size of the phase-retarding disc on the appearance under test of a true mirror are considered, and the conclusion is drawn that it is these effects, rather than considerations of sensitivity, which mainly determine the upper limit to the useful size of the disc.

---

§ 1. INTRODUCTION

AN account of the diffraction theory of the phase-contrast method of testing the figures of optical surfaces has been given in two interesting papers by its originator, F. Zernike (1934 a, b). However, the mathematical analysis in these papers takes a form which is not very well adapted to a quantitative discussion of the properties of the test. Moreover, it is for the most part restricted to surfaces whose errors are small compared with $\lambda/4\pi$, where λ is the wave-length of the light, that is to say, to surfaces which are already practically perfect.

In the present paper we re-develop the theory of the phase-contrast test and obtain formulae which predict the appearance under test of a mirror with arbitrary smooth errors of figure, not necessarily small compared with λ. These formulae are applied to the discussion of the appearance of a true mirror under·the test and the limit of sensitiveness of the test for local zonal errors. The analysis, like that of Zernike, is restricted throughout to the mathematically simplest case where the light-source (pinhole) is small compared with the Airy disc of the mirror under test and the light is monochromatic. In practice, the pinhole is usually comparable in size with the Airy disc and the light polychromatic, but a discussion of the extent to which the properties of the test are modified by these circumstances is beyond the scope of the present paper.

§ 2. THE PHASE-CONTRAST TEST

The notation and assumptions are similar to those adopted in a recent discussion of the Foucault test (Linfoot, 1946). For convenience they are recapitulated here.

All the cases where a convergent pencil is being null-tested can be covered by supposing that the wave-fronts originate at the surface of a nearly spherical mirror M which is being tested at its approximate centre of curvature. See figure 1. The results of the test are then interpreted in terms of errors of figure of the mirror M. We suppose that these errors of figure may amount to several wave-lengths, but that the error-slopes on the mirror, besides being free from discontinuities, are not so steep as to spread out the visible image to more than a moderate multiple, say 5 or 10, of the size of the Airy disc. Then it makes no appreciable difference if we suppose the wave to be leaving a true spherical

surface M_0, lying everywhere within a few wave-lengths of the surface M, and suppose that at the point (x, y, z) on M_0 the complex displacement is

$$E(x, y) = |E(x, y)| e^{\frac{2\pi i}{\lambda} \phi (x, y)} ;$$

thus $|E(x, y)|$ is the amplitude at the point (x, y) on M_0 and $\phi(x, y)$ the phase there; the variation in $\phi(x, y)$ expresses the distortion of the wave-fronts. Outside the boundary-circle C, lying in M_0 and corresponding to the edge of the mirror, we define $E(x, y)$ as zero.

After leaving the mirror, the light comes to a more or less imperfect focus in the neighbourhood of the centre of curvature O of M_0, at a distance s from its pole A; we suppose that s is large·compared with the diameter of the mirror. Just behind O, a viewing system L receives the light and forms an image M_0' of the surface M_0; we denote by C' the boundary of M_0'.
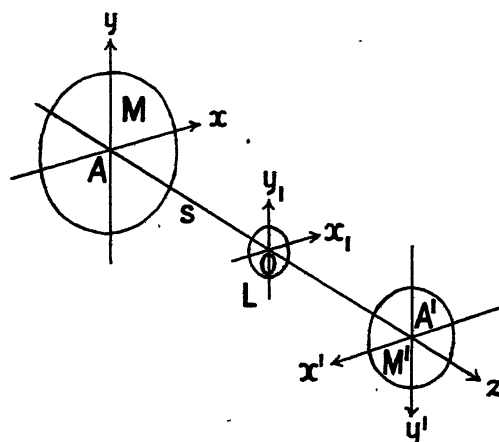
Figure 1.

At O is placed the phase-retarding Zernike disc, comparable in size with the Airy disc of the mirror, which renders variations of phase in the wave-fronts leaving M_0 visible as variations of intensity in the image-surface M_0'. x_1 and y_1, measured parallel to x and y, are coordinates in the space near O.

We define a system of coordinate-numbers (x', y') in the surface M_0' by assigning to the image P' of the point $P = (x, y, z)$ in M_0 the coordinate-numbers $x' = x$, $y' = y$. The intensity at P' is thus the same thing as the " intensity seen under the test " at P. To avoid unessential complications, we suppose that L is a perfect optical system.

By the usual arguments, based on an application of Huyghens' principle,[*] the complex displacement in the intermediate image-surface, which we take to be the sphere S of centre A passing through O, is

$$W(u, v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{ixu+iyv} E(x, y) dx\, dy, \qquad \ldots\ldots(2.1)$$

where

$$u = \frac{2\pi x_1}{\lambda s}, \qquad v = \frac{2\pi y_1}{\lambda s}, \qquad \ldots\ldots(2.2)$$

[*] See the paper already referred to for a fuller account of these arguments.

and the integral is only formally over an infinite domain, since the integrand vanishes outside the boundary-contour C.

If the phase-retarding disc were not present, the complex displacement $D(x', y')$ at the point (x', y') in the final image-surface would be obtained by integrating the product

$$\frac{1}{2\pi} e^{-i(x'u+y'v)} W(u, v)$$

over the area $u^2+v^2 \leq R^2$ of the (u, v)-plane which falls within the aperture of the viewing system L.* The effect of a disc occupying the region $u^2+v^2 \leq a^2$ of the intermediate image-surface and retarding by α radians the phase of the waves passing through it can be represented by changing $W(u, v)$ into $e^{-i\alpha}W(u, v)$ throughout this region and leaving it unaltered elsewhere. In saying this, we assume that the effect of the disc is merely to change the phase of those parts of the wave which impinge upon it, without influencing the neighbouring parts. This assumption is never strictly true, but it is a permissible approximation when the pencil under test is of small numerical aperture.†

The complex displacement at the point (x', y') of the final image-surface is then given, to a sufficient approximation, by the equation

$$D(x', y') = \frac{1}{2\pi} \iint\limits_{u^2+v^2 \leq R^2} e^{-iux'-ivy'} W(u, v)\, du\, dv$$

$$+ \frac{1}{2\pi}(e^{-i\alpha}-1) \iint\limits_{u^2+v^2 \leq a^2} e^{-iux'-ivy'} W(u, v)\, du\, dv. \qquad \ldots\ldots(2.3)$$

The first term of (2.3) is simply the complex displacement which results when the mirror is imaged through the viewing system, without the interposition of the Zernike disc, and the second term therefore describes the effect of the disc.

We next replace R by ∞ in the first term of (2.3). This amounts to neglecting the effects of the finite aperture of the viewing telescope. For mirrors with errors of the type under consideration, the resulting change in the predicted intensities is too small to be detected by the eye, except at points whose distance from the rim of the mirror is comparable with the resolving power of the viewing system. At these points the intensities depend critically on the aperture and optical aberrations of the viewing system, as well as on the errors of the mirror, and we therefore exclude them from the discussion. At the remaining points we obtain, on setting $R = \infty$ in (2.3), the equation

$$D(x', y') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-iux'-ivy'} W(u, v)\, du\, dv$$

$$+ \frac{1}{2\pi}(e^{-i\alpha}-1) \iint\limits_{u^2+v^2 \leq a^2} e^{-iux'-ivy'} W(u, v)\, du\, dv$$

$$= E(x', y') + \frac{1}{2\pi}(e^{-i\alpha}-1) \iint\limits_{u^2+v^2 \leq a^2} \bar{e}^{-iux'-ivy'} W(u, v)\, du\, dv \qquad \ldots\ldots(2.4)$$

* In the case of a $f/10$ pencil and a viewing telescope of 1-inch aperture the value of R is approximately 1300.

† For in this case the scale of the diffraction pattern in S is large compared with the wavelength λ. The radius of the first Airy dark ring is 3·83 in (u, v)-units.

for the complex displacement $D(x', y')$. The intensity seen under the test, at the point (x', y') on the mirror-surface or in the halo, is then measured by the quantity

$$I(x', y') = |D(x', y')|^2. \qquad \qquad \ldots \ldots (2.5)$$

§3. TRANSFORMATION OF (2.4): GENERAL FORMULAE

The second term in (2.4) describes the change in $D(x', y')$ caused by the presence of the Zernike disc. On substituting for $W(u, v)$ from (2.1) and remembering that $E(x, y)$ vanishes outside the part of M_0 corresponding to the mirror-surface, we can write this term in the form

$$\frac{1}{4\pi}(e^{-i\alpha} - 1) \iint_{u^2+v^2 \leq a^2} e^{-iux'-ivy'} \, du \, dv \iint_M e^{iux+ivy} E(x, y) \, dx \, dy$$

$$= \frac{1}{4\pi}(e^{-i\alpha} - 1) \iint_M E(x, y) \, dx \, dy \iint_{u^2+v^2 \leq a^2} e^{iu(x-x')+iv(y-y')} \, du \, dv. \qquad \ldots \ldots (3.1)$$

The inner integral

$$\iint_{u^2+v^2 \leq a^2} e^{iu(x-x')+iv(y-y')} \, du \, dv = 2\pi a^2 \frac{J_1(aZ)}{aZ},$$

where $Z = \sqrt{[(x-x')^2 + (y-y')^2]}$ and J_1 is the Bessel function of order 1. Hence, (2.4) can be written in the form

$$D(x', y') = E(x', y') + \frac{a^2}{2\pi}(e^{-i\alpha} - 1) \iint_M E(x, y) \frac{J_1(aZ)}{aZ} \, dx \, dy. \qquad \ldots \ldots (3.2)$$

It is possible to obtain (3.2) by arguments of a more physical character. The Zernike disc can be regarded as sending out a " virtual " supplementary wave (namely, one produced backwards in time) which interferes at the surface M_0 with the original wave $E(x', y')$. The resulting complex displacement $D(x', y')$ on M_0 determines the observed intensity $I(x', y') = |D(x', y')|^2$. Many of the properties of the Zernike test can be derived qualitatively by regarding the supplementary wave as a spherical wave originating at the centre of the Zernike disc,[*] but this simplified picture does not, of course, suffice for the derivation of (3.2). The idea can be carried through quantitatively, however, by applying the principle of superposition in the following way. Each element $dx \, dy$ of the mirror-surface sends out a spherical wave which is partly intercepted by the Zernike disc. The complex displacement which this wave causes in the surface of a sphere centred at the element $dx \, dy$ and passing through O (see figure 1) is $\frac{i}{\lambda s} e^{-\frac{2\pi i}{\lambda} s} E(x, y) \, dx \, dy$. The corresponding contribution to the virtual wave emitted by the Zernike disc is represented, to a sufficient approximation, by a complex displacement $\frac{i}{\lambda s} e^{-\frac{2\pi i}{\lambda} s} (e^{-i\alpha} - 1) E(x, y) \, dx \, dy$ over the region of this sphere which most nearly coincides with the area of the disc, namely the circle of arcual centre O and of radius $\frac{\lambda s}{2\pi} a$. Now, by the classical theory of the Airy

---

* See C. R. Burch (1934).

disc, a real wave-front filling this circle and with complex displacement-function equal to 1 over its area would produce at the point (x', y') on M_0 a complex displacement

$$\frac{2\pi i}{\lambda s} e^{-\frac{2\pi i}{\lambda} s} \left(\frac{\lambda s}{2\pi} a\right)^2 \frac{J_1(aZ)}{aZ},$$

where $Z = \sqrt{[(x-x')^2 + (y-y')^2]}$. In the time-reversed "virtual" wave, the factor $\frac{2\pi i}{\lambda s} e^{-\frac{2\pi i}{\lambda} s}$ becomes $-\frac{2\pi i}{\lambda s} e^{\frac{2\pi i}{\lambda} s}$, while the initial complex displacement-function is not 1 but $\frac{i}{\lambda s} e^{-\frac{2\pi i}{\lambda} s} (e^{-i\alpha} - 1)E(x, y)dx\,dy$. If follows that the contribution from the element $dx\,dy$ to the virtual supplementary wave sent out by the Zernike disc is such as to produce at a point (x', y') in the surface M_0 the complex displacement

$$\frac{i}{\lambda s} e^{-\frac{2\pi i}{\lambda} s} (e^{-i\alpha} - 1)E(x, y)dx\,dy \,.\, -\frac{2\pi i}{\lambda s} e^{\frac{2\pi i}{\lambda} s} \left(\frac{\lambda s}{2\pi} a\right)^2 \frac{J_1(aZ)}{aZ}$$

$$= \frac{a^2}{2\pi} (e^{-i\alpha} - 1)E(x, y) \frac{J_1(aZ)}{aZ}\, dx\,dy.$$

Integrating over the elements $dx\,dy$, we obtain

$$\frac{a^2}{2\pi}(e^{-i\alpha} - 1) \iint\limits_M E(x, y)\, \frac{J_1(aZ)}{aZ}\, dx\,dy$$

as the total complex displacement at (x', y') due to the virtual wave from the Zernike disc, and (3.2) follows at once.

Evidently this argument can be applied to mirrors of arbitrary edge-contour and variable reflecting power; by taking α complex in the factor $(e^{-i\alpha} - 1)$ we can also cover the case where the Zernike disc absorbs a known fraction of the light which passes through it. But since practical interest centres mainly on the circular mirror of uniform reflecting power, tested with a disc whose absorption and back-reflection can be disregarded, the discussion is here restricted to this case.

To calculate the intensities seen under the test on a mirror whose errors of figure correspond to the presence in the image of small amounts of the classical aberrations, we expand the function $E(x, y)$ as a series

$$E(x, y) = E(r \cos \phi, r \sin \phi) = \sum_{m=-\infty}^{\infty} E_m(r)e^{im\phi}$$

over the mirror-surface $0 \leq r \leq 1$, $0 \leq \phi \leq 2\pi$. The integral (3.1) can then be written

$$\frac{1}{4\pi^2} (e^{-i\alpha} - 1) \int_0^1 r\,dr \int_0^a \rho\,d\rho \int_0^{2\pi} E(r \cos \phi, r \sin \phi)d\phi \int_0^{2\pi} e^{ir\rho \cos(\phi - \chi) - ir'\rho \cos(\phi' - \chi)}d\chi,$$

where $\qquad\qquad x + iy = re^{i\phi}, \; x' + iy' = r'e^{i\phi'}, \; u + iv = \rho e^{i\chi}$

$$= \frac{1}{2\pi} (e^{-i\alpha} - 1) \int_0^a \rho\,d\rho \int_0^1 r\,dr \int_0^{2\pi} E(r \cos \phi, r \sin \phi)J_0(\rho Z)d\phi.$$

where
$$Z = \sqrt{[(x-x')^2 + (y-y')^2]}$$

$$= \frac{1}{2\pi}(e^{-i\alpha} - 1)\sum_m \int_0^a \rho \, d\rho \int_0^1 E_m(r)r \, dr \int_0^{2\pi} e^{im\phi}J_0(\rho Z)d\phi$$

$$= (e^{-i\alpha} - 1)\sum_m \int_0^a \rho \, d\rho \int_0^1 E_m(r)J_{|m|}(\rho r)J_{|m|}(\rho r')r \, dr.$$

(3.2) thus becomes

$$D(x',y') = E(x',y') + (e^{-i\alpha}-1)\sum_{m=-\infty}^{\infty} e^{im\phi'}\int_0^1 E_m(r)r \, dr \int_0^a J_{|m|}(\rho r)J_{|m|}(\rho r')\rho \, d\rho$$

$$= E(x',y') + a^2(e^{-i\alpha}-1)\sum_{m=-\infty}^{\infty} e^{im\phi'}\int_0^1 E_m(r)K_{|m|}(ar, ar')r \, dr,$$

$$\ldots\ldots(3.3)$$

where $K_m{}'$ is defined by the equations

$$K_m(u,v) = \frac{1}{u^2-v^2}[uJ_{m+1}(u)J_m(v) - vJ_{m+1}(v)J_m(u)] \quad (u \neq v)$$

$$= -\frac{1}{2u}[uJ_{m+1}(u)J_m{}'(u) - uJ_m(u)J'_{m+1}(u) - J_m(u)J_{m+1}(u)] \quad (u=v).$$

$$\ldots\ldots(3.4)$$

More particularly, following Zernike, we may expand $E(x, y)$ as a series of polynomials in x and y, orthogonal over the unit circle, by writing

$$E(x, y) = \sum_{m, n} a_{m, n}R_n{}^{|m|}(r)e^{im\phi}, \qquad \ldots\ldots(3.5)$$

where, in the summation, $-n \leq m \leq n$ and $m-n$ is even, while $R_n{}^m(r)$ is the polynomial

$$(-1)^{\frac{n-m}{2}}\left(\frac{\frac{1}{2}(n+m)}{m}\right)r^m F\left(\frac{n}{2} + \frac{m}{2} + 1, -\frac{n}{2} + \frac{m}{2}, m+1, r^2\right).$$

The equation for $D(x', y')$ then takes the form .

$$D(x', y') = E(x', y') + (e^{-i\alpha}-1)\Sigma a_{mn}e^{im\phi'}\int_0^a J_{|m|}(\rho r')\rho \, d\rho \int_0^1 R_n^{|m|}(r)J_{|m|}(\rho r)r \, dr$$

$$= E(x', y') + (e^{-i\alpha}-1)\Sigma(-1)^{\frac{n-m}{2}} a_{mn}e^{im\phi'}\int_0^a J_{|m|}(\rho r')J_{n+1}(\rho)d\rho$$

$$= E(x', y') + (e^{-i\alpha}-1)\Sigma(-1)^{\frac{n-m}{2}} a_{mn}e^{im\phi i} I_{m, n}(r', a), \qquad \ldots\ldots(3.6)$$

where

$$I_{mn}(r', a) = \int_0^a J_{|m|}(\rho r')J_{n+1}(\rho)d\rho.$$

The true mirror

In the case of a true mirror tested with the Zernike disc at focus, $a_{00} = 1$ and the remaining coefficients a_{mn} are zero; thus

$$D(x', y') = D(r') = \frac{1}{0} + (e^{-i\alpha}-1)I(r', a), \qquad \ldots\ldots(3.7)$$

where *

$$I(r', a) = I_{00}(r', a) = \int_0^a J_0(r'\rho)J_1(\rho)d\rho$$

$$= (1 - J_0(a))\frac{2J_1(ar')}{ar'} + (1 - J_0(a) - 2J_2(a))\frac{6J_3(ar')}{ar'}$$

$$+ (1 - J_0(a) - 2J_2(a) - 2J_4(a))\frac{10J_5(ar')}{ar'} + \dots, \qquad \dots\dots(3.8)$$

and in (3.7) the upper or lower alternative is to be taken according as $x'^2 + y'^2 \lessgtr 1$.

Figure 2a, which is similar to one given by Zernike (1934a), shows the intensities $I_0 = I_0(r') = |D(r')|^2$ seen on the mirror and in the halo with quarter-wave discs $\left(\alpha = \frac{\pi}{2}\right)$ of radii $a = 2\cdot5$, $a = 2\cdot0$ and $a = 1\cdot6$. It will be seen that in the first two cases the mirror appears brighter in the centre. In the practical use of the test, the focal setting of the Zernike disc is empirically chosen so as to make
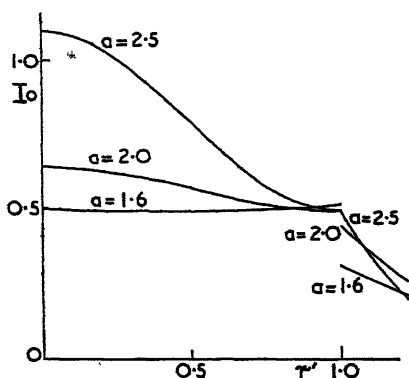
Figure 2a. True mirror; intensities with phase-retarding disc at focus.
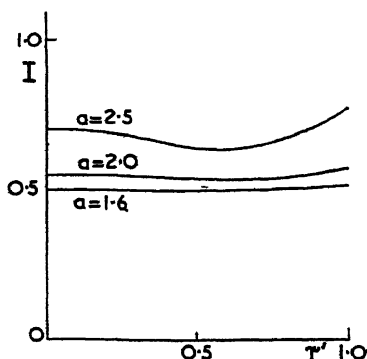
Figure 2b. True mirror; with phase-retarding disc at preferred focal setting.

the intensities seen on the mirror surface as uniform as possible. It follows that in the first two cases the disc will be set, not at focus, but a little inside it. Figure 2b† shows the intensities seen on a true mirror at the focal settings which would actually be used in practice for these three values of a; the amount of defocusing is about $\frac{1}{12}$ fringe in the first case and $\frac{1}{50}$ fringe in the second; in the third case, the disc is at focus. As the size of the Zernike disc is increased, the residual inequalities in the intensity at selected focus become more serious, and it may be inferred from figure 2b that, in the case where the dimensions of the light-source can be neglected, the radius of a quarter-wave Zernike disc must

* Zernike (1934a), p. 703.
† Calculated from (3.10); see next page.

not exceed about two-thirds of that of the Airy disc if a true mirror under the test is to appear uniformly illuminated.†

Zonal errors

A more convenient expression than (3.6) for $D(x', y')$ can be obtained when the surface under test suffers only from zonal errors, that is, when $E(x, y) = E(r)$ is a function of r alone. In this case $D(x', y') = D(r')$ is likewise radially symmetrical and is given by the equation

$$D(r') = E(r') + (e^{-i\alpha} - 1) \int_0^1 E(r) r \, dr \int_0^a J_0(\rho r) J_0(\rho r') \rho \, d\rho \qquad \ldots\ldots(3.9)$$

$$= E(r') + a^2(e^{-i\alpha} - 1) \int_0^1 E(r) K_0(ar, ar') r \, dr$$

$$= E(r') + a^2(e^{-i\alpha} - 1) \int_0^1 E(r) \frac{ar J_1(ar) J_0(ar') - ar' J_1(ar') J_0(ar)}{(ar)^2 - (ar')^2} r \, dr.$$

$$\ldots\ldots(3.10)$$

Equation (3.10) reduces the problem of calculating the appearance under test of a mirror suffering from zonal errors of arbitrary form to the carrying out of a pair of numerical integrations for each value of r'.

§4. SENSITIVITY OF THE TEST FOR ZONAL ERRORS

We use for shortness the operational notation

$$Z_\alpha E(r') = \int_0^1 E(r) r \, dr \int_0^a J_0(\rho r) J_0(\rho r') \rho \, d\rho. \qquad \ldots\ldots(4.1)$$

Thus $Z_\infty E(r') = E(r')$, by the Fourier-Hankel inversion formula. (3.9) can now be written

$$D(r') = E(r') - (1 - e^{-i\alpha}) Z_\alpha E(r'). \qquad \ldots\ldots(4.2)$$

In the case of a true mirror tested with Zernike disc at focus, the wave-function is

$$E_0(r) = 1 \ (r \leq 1)$$
$$= 0 \ (r > 1),$$

and $Z_\alpha E_0(r') = I(r', a) = I_a$ can be calculated from (3.8). When the zonal errors are small, for example when they do not exceed 1/20 fringe, it is permissible to write

$$E(r) = 1 + iE^*(r) \ (r \leq 1)$$
$$= 0 \qquad\qquad (r > 1),$$

where $E^*(r)$ is real, and to neglect $(E^*)^2$. (4.2) then gives, if $r' < 1$,

$$D(r') = 1 + iE^*(r') - (1 - \cos\alpha + i\sin\alpha)(I_a + iZ_\alpha E^*(r'))$$
$$= 1 - (1 - \cos\alpha)I_a + \sin\alpha Z_\alpha E^*(r')$$
$$\quad + i[E^*(r') - I_a \sin\alpha - (1 - \cos\alpha)Z_\alpha E^*(r')]$$

and the intensity

$$I = I(r') = |D(r')|^2 = 1 - 4\sin^2\frac{\alpha}{2} I_a(1 - I_a) - 2\sin\alpha[E^*(r')I_a - Z_\alpha E^*(r')]$$

† Zernike (1934 b, p. 383) recommends the use of a disc between one-half and two-thirds of the size of the Airy disc for detecting astigmatism, and asserts that for detecting spherical aberration a disc more than twice as large as the Airy disc can be used. But his method of arriving at these estimates is not free from objection, since in the paper referred to no account is taken of the variation of intensity over a true mirror at preferred focus, on which chiefly depends, according to the present analysis, the upper limit to the useful size of the disc when small errors are under examination.

on discarding terms in $(E^*)^2$. That is to say

$$I - I_0 = -2 \sin \alpha [E^*(r')I_a - Z_a E^*(r')], \qquad \dots\dots(4.3)$$

where
$$I_0 = 1 - 4 \sin^2 \tfrac{\alpha}{2} I_a (1 - I_a) \qquad \dots\dots(4.4)$$

is the intensity for a true mirror and

$$I_a = I(r', a) = \int_0^a J_0(\rho r') J_1(\rho) d\rho$$

can be calculated from (3.8).

In the case $\alpha = \tfrac{\pi}{2}$ of a quarter-wave retarding disc, (4.3) agrees with the corrected form of a result of Zernike.† When $r' > 1$, i.e. in the halo, (4.2) gives

$$D(r') = -(1 - e^{-i\alpha}) Z_a E(r')$$
$$= -(1 - \cos \alpha + i \sin \alpha)(I_a + i Z_a E^*(r')),$$
$$I(r') = |D(r')|^2 = 4 \sin^2 \tfrac{\alpha}{2} I_a^2. \qquad \dots\dots(4.5)$$

Thus the appearance of the halo is not visibly changed by small zonal errors on the mirror. It is easy to show from (3.2) that the same conclusion follows when the errors are not restricted to be radially symmetrical.

From (4.3) we can predict in a general way the appearance under test of a shallow high (or low) zone on an otherwise true mirror. To fix ideas, let $\alpha = \tfrac{\pi}{2}$, $a = 2 \cdot 0$ and suppose for the present that the disc is at focus. Then (4.3) becomes

$$I(r') = I_0(r') - 2 I_a E^*(r') + 2 Z_a E^*(r').$$

On the right-hand side of this equation, $I_0(r')$ runs from 0·65 to 0·5 as r' increases from 0 to 1 (see figure 2 (a)), while $2 I_a$ runs from 1·55 to 0·95. The second term, $-2 I_a E^*(r')$, therefore contributes an intensity-decrease approximately proportional, at each point of the zone, to the height-excess of the surface at that point, and approaching 50% of the value of I_0 at a point where $E^*(r') = 1/4$, i.e. where the surface is $\tfrac{1}{25}$ fringe too high. The third term, $2 Z_a E^*(r')$, describes the manner in which the light which has disappeared from the high zone is re-distributed, fairly uniformly, over the whole mirror; as (4.5) shows, practically none of this light is sent into the halo. This redistributed light reduces the contrast with which the zone is shown up under the test, and its effect becomes more noticeable as the area of the zone increases.

A complication, the effect of which on the sensitivity of the test can be foreseen to be small without exact calculation, is that in practice the Zernike disc will not be set at focus but, in the present case, about $\tfrac{1}{50}$ fringe inside focus. It therefore seems safe to conclude that a fairly narrow zonal error of $\tfrac{1}{20}$ fringe should be easily visible on an otherwise true mirror and hence that the sensitivity of the test for errors of this type is of the same order as that of the Foucault test. As a check on this estimate, the intensity-distributions in a number of typical cases were computed from (3.10). A selection from the results is given in figure 3,

† Zernike, 1934, p. 704, line 11, where the effect of the term in $Z_a E^*(r')$ is not taken into consideration ; there is also a mistake in sign.

which shows the appearance of high and low zones of $\frac{1}{20}$ fringe under the Zernike test at the preferred focus, for three different values $a=2\cdot5$, $a=2\cdot0$ and $a=1\cdot6$ of the disc-radius, and under the Foucault test with the knife-edge centrally set.*·

From figure 3 it appears that, for the detection of an error of this type, there is no appreciable difference in sensitivity between quarter-wave discs of radii $a=2\cdot0$ and $a=2\cdot5$, but that, with the larger disc, less light is diverted into the halo, and so lost for testing purposes. The sensitiveness is perceptibly decreased,
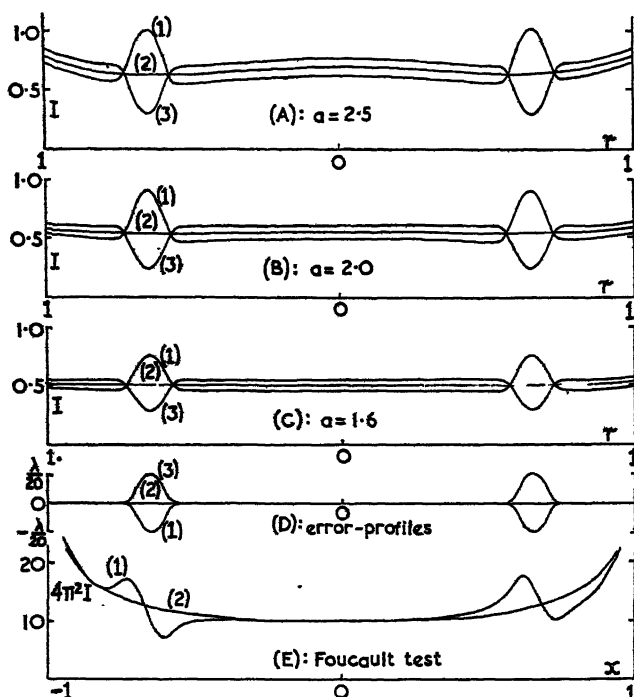
Figure 3.   (A)-(C) Intensities under Zernike test with quarter-wave retarding disc, of radii $a=2\cdot5$, $2\cdot0$ and $1\cdot6$, at preferred focus: (1) low zone, (2) true mirror, (3) high zone. (D) Error profiles. (E) Intensities under Foucault test, knife-edge central ; (3) is obtained on reversing (1) from left to right.

however, if the value of a is reduced to $1\cdot6$. We conclude that, in testing local zonal errors, the sensitiveness of the test is nearly independent of the disc-size over the range $2\cdot0 \leqq a \leqq 2\cdot5$, and the upper limit to the useful size of the Zernike disc is accordingly determined by the consideration that a true mirror must not show too great inequalities in brightness when tested at the preferred focal setting.

* The Foucault intensities were calculated from the formula

$$4\pi^2 I(x', 0) = \left| \pi E(x', 0) + i \int_{-1}^{1} \frac{E(t, 0)}{t - x'} \, dt \right|^2 \quad (-1 < x' < 1).$$

REFERENCES

Burch, C. R., 1934.   *Mon. Not. R. Astr. Soc.*, **94**, 384.
Linfoot, E. H., 1946.   *Proc. Roy. Soc.*, A, **186**, 72.
Zernike, F., 1934 a.   *Physica*, **1**, 689; 1934 b.   *Mon. Not. R. Astr. Soc.*, **94**, 377.

THE EFFECT OF THE ANGLE OF INCIDENCE OF THE EXPOSING LIGHT RAYS UPON THE RESOLVING POWERS OF PHOTOGRAPHIC MATERIALS

By J. M. GREGORY

ABSTRACT. It has been supposed that the angle of incidence of the exposing light would influence the apparent resolving power of a photographic material. Experiments have been carried out with a resolving power test object in contact with the photographic material, using illumination at different angles of incidence. The effect is almost indiscernible until angles of 40° are reached, at which it is of the order of 5 to 10 % reduction in resolving power with a medium-speed film. With fast film for aerial photography, the effect is barely observable. It is concluded that the drop in resolving power, observed in air camera negatives at points off axis, is not to any appreciable extent due to the above effect.

§ 1. INTRODUCTION

EXPERIMENTS have shown that the photographic resolution in camera negatives falls off at the edges of the field when the lens is focused on the film for the central part of the field. The reasons for this effect became of interest under a programme of research carried out for the Air Photography Research Committee of the Ministry of Aircraft Production, designed to provide the information necessary to advise on the improvement of the lens-film combination as used in air photography.

Selwyn and Tearle (1946) have discussed the effect in detail. It was suggested during discussions that the loss in resolution was due to the fact that, off the optic axis of the lens, the light enters the photographic emulsion obliquely whilst the processed film is, necessarily, viewed from the perpendicular. The effect expected is illustrated in figure 1 (b) in a highly idealized form. A screen, AA, having two apertures as shown, is illuminated by parallel light such that the photographic material is exposed at an angle to the incident beam. The shaded areas in the sensitive layer below it show the effect of the light on the layer as revealed by the subsequent processing. Examination from the perpendicular direction will no longer separate the two exposed areas, as would be the case when the exposing rays are normal to the sensitive layer (figure 1 (a)). This possibility was considered by lens-designers at one time to be an important factor setting a limit to off-axis resolution quite apart from considerations of the off-axis lens aberrations.

Dr. Stevens of these Laboratories has been able to demonstrate that effects of this nature can occur with a very transparent emulsion such as that of " Kodak Maximum Resolution Plates ", but the work described here shows that there is no appreciable effect of this sort with negative materials of the type used in aerial photography.

It does not seem possible to devise a method for the direct determination of the effect of angle of incidence of the light rays upon photographic resolution which can give completely unequivocal results. Two methods, broadly speaking, are available. They may be termed the " Lens " method and the " Contact Printing "

method respectively.   The former method makes use of a very well corrected lens to throw an image of the test object upon the photographic material.   It suffers from the disadvantage in the present application that as the photographic material is tilted to obtain the various angles of incidence required, the various parts of the test-object image not only go out of focus, but the image of the test object is also distorted.   The latter method, which is the one used in this work, makes use of a special test object which is placed in good contact with the material under test, and exposed in much the same way as a contact print made from a photographic negative.   The object is special in the sense that it has to be made much smaller than the test objects used for the lens method.   In consequence, the photographic material used for making the test object must have the very highest possible resolution to record with sufficient perfection the minute images concerned.
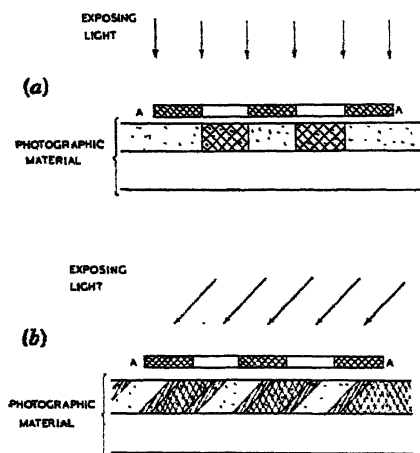
Figure 1.   Diagrams to show the expected effect of obliquity of incident light and the thickness of test object on resolution.
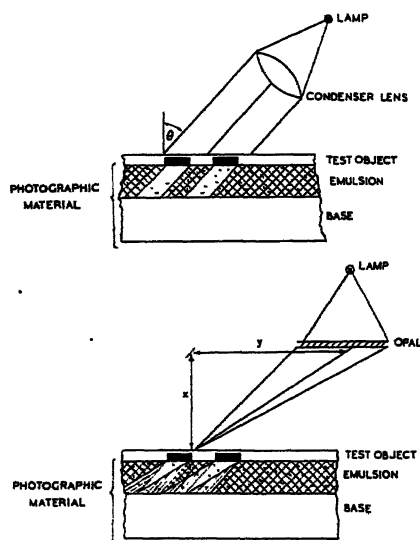
Figure 2.   Experimental arrangement.

"Kodak" Maximum Resolution plates have been found suitable in this respect. The test objects were made with the microphotographic apparatus described by Stevens (1944).

The disadvantage of this method is that the effective contrast of the test object changes with angle of incidence of the exposing light ray.   The ratio of the path length, L, to the test-object thickness, t, for an incident angle θ, is given by

$$L/t = \left[1 - \frac{1}{\mu^2} \sin^2\theta \right]^{-\frac{1}{2}},$$

μ being the refractive index (air to test-object layer).

Since the ratio is independent of the thickness, the effect cannot be eliminated experimentally by reducing the layer thickness.   It can, however, be allowed for, and this will be dealt with later.   Another difficulty, connected with the finite thickness of the opaque layer, is that the effective width of the various parts of the test object changes with the angle of incidence.   This can be seen in figure 1 (*b*).

The difficulty can be overcome by making the opaque layer thin enough, since the effect depends upon the ratio of the layer thickness to the separation of the opaque parts of the individual charts of the test objects.   For a given

separation, therefore, the ratio can be made as small as we like by making the layer thickness sufficiently small.

For resolutions of 30 lines per millimetre (the order of resolution of air photographic materials) the separation between the bars of the test-object image is 17μ. Measurements, on the particular test object used in this work, using a microscope with an oil-immersion objective, show that the images are 3 to 4μ thick; at 45° angle of incidence this would mean geometrically a reduction of the apparent separation by 3 to 4μ. Thus we might expect a purely geometrical reduction of the resolution by some 20% at 45°. This argument, however, has omitted consideration of the contrast profile of the test object as the angle of incidence increases. It can be seen that the path length of a light ray inclined to the normal going through the test object decreases as it approaches the edge. The effect of decreasing apparent separation, as the angle of incidence increases, must be largely compensated for by the parallel reduction in test-object contrast at the edges.

We have so far considered that the inclination of the ray takes place in a plane which is at right angles to the direction of the "bars" of the test chart. The light-scattering effects in the emulsion then take place across the area of material corresponding to the space between the bars of the test chart. We shall, however, be generally interested in the effect of inclination of the light ray regardless of the direction of the ray relative to the test-chart bars, for, in practice, details in the subject have a shape requiring good resolution in all directions.

The illumination of a point on a film given by a lens comes from all parts of the lens and forms a conical bundle of light rays pointing at the image point. One might expect, if resolution is at all effected by the angle of incidence of a light ray, that such a conical bundle of light rays could have a considerably greater effect than a beam of parallel light incident at the same angle as the axis of the cone. This point was therefore investigated experimentally by using an illuminated opal disc as a source of light for the contact exposures.

§2. EXPERIMENTAL

The diagrams of figure 2 show the two kinds of experiments which were made. In the first series of experiments, the pressure frame, holding the test object and test material clamped in good contact, was placed in the beam of light from a 5-inch diameter condenser lens arranged at such a distance from a 70-watt exciter lamp (which has a very compact filament) that the beam consisted of practically parallel light. By tilting the holder relative to the beam, a set of exposures was made at four different angles of incidence, namely 0°, 20°, 40° and 60° to the normal. In the second series of experiments, an evenly illuminated opal disc, of 6-inch diameter, was used as the source of illumination. It was placed parallel to the test object and distant 38 inches from it. The same four angles of incidence were obtained by shifting the frame parallel to itself by appropriate amounts, y, maintaining the distance, x, at 38 inches. Thus when disc and test object are opposite each other, the aperture of an equivalent lens would be f/6·3. This was chosen because it was one of the most common apertures in operational use for air-reconnaissance photography. In what follows, the first of these arrangements is referred to as yielding "parallel" illumination and the second as "diffuse".

The test object employed consisted of a row of units, each containing a range of sizes of Cobb charts. The unit is shown in figure 3. The Cobb charts are arranged in groups of a dozen all the same size but alternately at right angles to each. other. The groups with the largest Cobb charts are placed on the outer edges of the unit. The size increments are 10% from group to group. By placing a neutral-density step wedge in front of the row of units a range of exposures could be obtained at the same time.
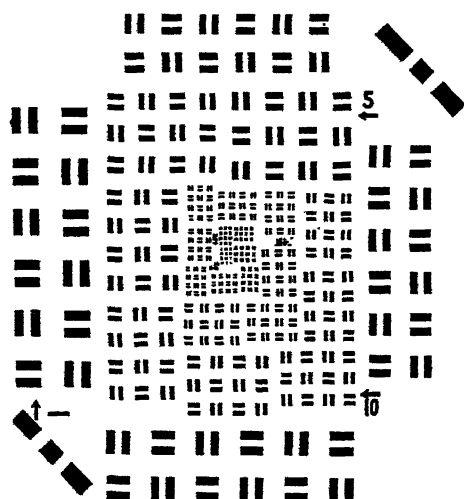
Figure 3.   Test-object unit.
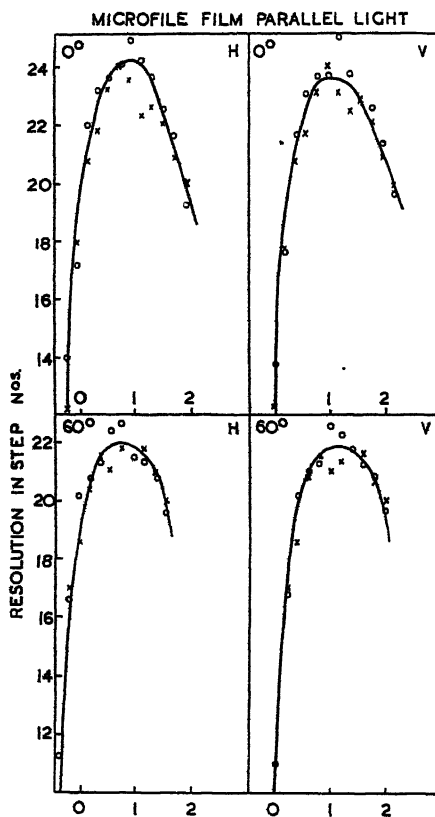
Figure 4.   Typical resolution-exposure curves.

The design of the groups in alternate charts at right angles means that the resolution in two directions at right angles can be determined at the same time, by recording the results for the two directions separately. This means that some estimate of the general effect, regardless of the ray direction relative to the chart bar, can be obtained by averaging these two results.

The contrast of the test object is moderately low, there being a density of 0·38 between the light and dark lines of a chart, i. e. the light lines have a brightness of approximately $2\frac{1}{2}$ times that of the dark lines.

Two films of widely differing light-scattering properties were tested in this way. They were "Kodak Aero Pan Super-XX", a very fast photographic material used extensively for air photography, and "Micro-File" Panchromatic, a slow fine-grained material. Two strips of each film were exposed at each angle of incidence. The average thickness of the "Aero Super-XX" emulsion layer was about 25 μ and of the "Micro-File" Panchromatic about 15μ.

All strips of each film were developed together in a machine giving high agitation of the developer. The " Super-XX" films were developed in Kodak developer D-19b for 8 minutes, and the "Micro-File" in Kodak developer D-16 for 4 minutes.

The processed strips were examined under a microscope at constant field brightness. Owing to the granular structure of photographic materials there is no sharp demarcation at which one group is resolved and the next smaller group is not resolved. Instead there is a fairly steady decrease in the number of individual charts resolved per group. The group at which 50% of the individual charts were resolved was therefore noted. A graph of exposure step *versus* group number was then drawn for each material and angle of incidence, the results for the repetition of each angle being plotted together.

§3. RESULTS

An example showing the resolution curve for "Micro-File" film at an angle of incidence of 60° with parallel light compared with the 0° curves is given in figure 4. The results for the two sets of test charts at right angles to each other are given separately under the arbitrary headings *vertical* (V) and *horizontal* resolutions (H), the former corresponding to the case when the ray inclination is in a plane at right angles to the chart bars. The *resolving power* of a material is defined as the maximum value of resolution on the resolution-log exposure curve. The average density at which the resolution-log exposure curve was a maximum was $D_{max} = 1\cdot0$ for the "Aero Super-XX" film and $D_{max} = 1\cdot3$ for the "Micro-File" panchromatic.

The resolving powers in the horizontal and vertical directions were all so nearly alike that they were averaged and the averages were then plotted as a function of the angle of incidence (figure 5 (a)).

§4. DISCUSSION OF RESULTS

The most noticeable features of the results were the lack of difference between the horizontal and vertical resolving powers and, to a less extent, the lack of an angle effect. It is only on reaching the extreme angle of 60° that a drop of resolution is encountered with the least scattering type of film, viz. "Micro-File". When the illumination is "diffuse" only the "Micro-File" results are influenced.

Before we can draw conclusions from these data, the effect of the test-object density with angle of incidence must be considered. Using the formula developed earlier for the increase in path length with angle of incidence, the percentage change in density may be computed. This is plotted in figure 5 (b) as a dotted line. Information on how such changes affect the resolution is meagre or inappropriate. The most appropriate data to hand are those of W. Romer (private communication) who used a well corrected quartz-fluorite apochromat by Hilger to throw the image of a series of test objects of various contrasts on to the same two photographic materials used in the present work. He found that in the region of test-object densities of 0-4, a 10% increase in density gave a 5% increase in resolution for "Micro-File" film and an 8% increase in resolution for "Super-XX". We may suppose with fair confidence that under the conditions of Romer's experiments, the effects found were mainly due to the material and that the lens did not interfere unduly with the results. By making proportionate allowances in the resolution figures for the change of contrast with angle of incidence (which applies equally to

horizontal and vertical resolution components) we obtain figure 5 (*b*). We see that from the practical point of view the corrections are very small indeed.

Turning back to the discussion of the results then, we see that there is a small effect of angle of incidence on resolution, which is of little importance until very oblique angles are encountered. The effects are greater for "Micro-File" film both as regards the angle of incidence and in regard to the use of the diffuse illumination. This is readily explained in terms of the much smaller scattering of light in this material. The greater the light scattering in an emulsion the less the effect of angle may be expected to be. In the "Super-XX" layer the scattering is so great that the original angle of incidence of the light, or the question whether the incidence light is diffuse or parallel, hardly matters. An interesting conclusion may be drawn from the fact that the "Micro-File" film resolutions at 60° are the same for both diffuse and parallel light. The "diffuse" illuminator is such that

ANGLE OF INCIDENCE
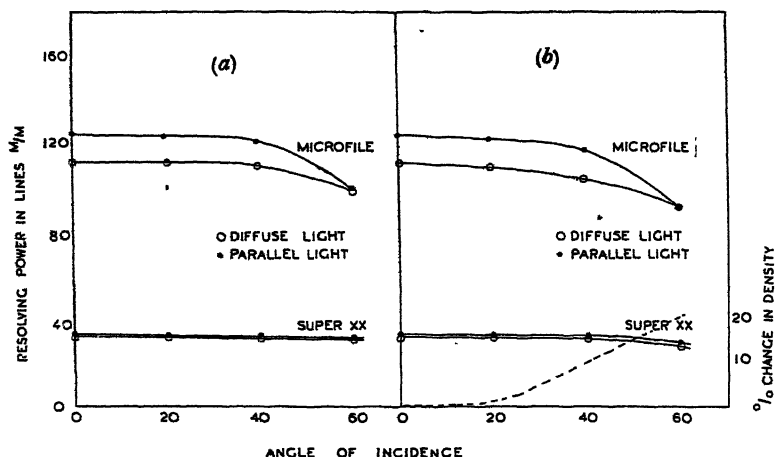
Figure 5. Graphs showing the effect of obliquity upon resolving power both before and after allowance is made for change in test-object density with obliquity. The percentage change in test object density with obliquity is shown by the dotted line.

the semi-angle of the illuminating cone gets progressively smaller as the angle of incidence increases, for the solid angle subtended by the opal disc at the test object decreases with increasing angle of incidence in the present set-up. This is due partly to the change in the projected size of the disc and partly to the increased distance from the test objects. Evidently, at 60°, the angle of the cone is so small that, as far as the "Micro-File" film is concerned, it is equivalent to "parallel" light.

We may conclude that there is an effect of angle to incidence of the exposing light upon the resolution of materials of the type used in air photography but that it is far too small to be of any importance in comparison with the drop in resolving power found when used in an air camera.

ACKNOWLEDGMENTS

REFERENCES

SELWYN E. W. H. and TEARLE, J. L., 1946. *Proc. Phys. Soc.*, **58**, 493.
STEVENS, G. W. W., 1944. *Photogr. J.*, **81**, 108.

DISCUSSION

on papers by R. A. Scott entitled
  (i) " The absorption of sound in a homogeneous porous medium."
  (ii) " An apparatus for accurate measurement of the acoustic impedance of sound-absorbing materials."
  (iii) " The propagation of sound between walls of porous material."
(This volume, pp. 165, 253 and 358.)

Dr. A. J. King. It may be helpful if I indicate briefly the circumstances which necessitated undertaking the three investigations described by the author in his group of papers. At the beginning of the war, we had in progress, in the Research Department of the company with which the author and I are connected, an investigation into the attenuation of sound in ducts such as are used in air-conditioning systems. This was to enable us to design ducts which would serve hotel bedrooms and ship's cabins without causing nuisance from noises due to fans and air turbulence. Measurements of sound attenuation were made in an experimental duct with various linings of porous material and an attempt made to correlate the results with the values predicted by the more or less empirical theories based on absorption coefficients put forward at that time. It was apparent that these theories were inadequate as they gave only a rough guide to the attenuation to be expected in any given case. When Morse published his theory it appeared to be much more comprehensive, but complete experimental verification was lacking.

At this point it became necessary to design a large engine-testing plant for war purposes, capable of testing jet engines at night without disturbing the neighbourhood. The magnitude and importance of the undertaking made it essential that the data used for design purposes should be reliable. Measurements were made of the intensity-frequency spectrum of the noise of an engine and of the ambient noise at night, thus enabling us to deduce the desirable attenuation-frequency characteristic of the silencing system. The existing investigation was therefore broadened, involving an extension of our acoustics laboratory to accommodate a more suitable duct and the development of apparatus, described in one of the papers, for the study of acoustic impedance. The agreement then observed between the experimental attenuations and those predicted by Morse's theory was much more satisfactory than in the case of the earlier theories, and so justified proceeding with the building of the test house. However, it was apparent that Morse's theory is incomplete, as it does not allow for the passage of sound along the material lining the duct. It was therefore necessary to study the mechanism of sound transmission in porous materials, and the author has shown that wave motion in them is similar to that in air. Unfortunately, the theory which the author has given, taking into account the passage of sound in the lining, does not lead to simple computation. However, he has shown that the difference between Morse's theory and the experimental results is in the direction indicated by his more complete theory, viz., that where Morse's theory predicts a high attenuation in a duct lined with loosely packed porous material, transmission along the lining is appreciable, resulting in experimental results lower than the predicted values. Where the predicted attenuation is lower, and the porous material denser, the difference is reduced. Thus, all three investigations were necessary for a complete appreciation of the problem.

Dr. G. H. Aston. I wish to ask to what extent the calculated attenuations, based on Dr. Scott's theory, for ducts with air spaces as used in practice have been compared with measured attenuations, and how far there is agreement.

Author's reply. In reply to Dr. Aston, the full range of ducts of interest in practice is very extensive and includes, at the one extreme, ducts which are split into channels of an inch or so in width by layers of absorbing material, and at the other extreme, ducts a few feet wide and lined with a few inches' depth of absorbing material. The first extreme lies close to the case referred to in figure 3 on p. 365. Thus, an examination of equation (15) on p. 362 shows that the attenuation coefficient calculated from my theory lies very close

to that characteristic of propagation of sound of the same frequency in material of the
lining in bulk. The experimental evidence put forward in figure 3 is in agreement with
this deduction. At the other extreme of very wide ducts, the attenuation is usually very
small (i.e. a few db./ft., and often much less). Comparison of equations \15) and (25)
shows that the solutions of my equation agree under this condition substantially with
those of Morse. In the course of a comprehensive investigation of the performance of
the ducts, we have shown in our laboratories that for wide duct-spacings there is good
agreement between measured attenuations and those calculated for the first propagational
mode from Morse's (and therefore from my own) theory. We hope to publish a full
account of this work in the near future.

There remains, however, the ducts for which the width of air-space and depth of lining
are intermediate between the above extremes. Over this range we have considerable
evidence of practical performance but practically no corresponding results calculated
from my theory. The reason for the deficiency is that numerical computation of solutions
of (15) in the general case are difficult. Prof. D. R. Hartree has suggested, in a very
helpful discussion of the problem with the author, a manner in which the differential
analyser might conveniently be used for the solution of the equations (7) and (9), and we
hope to pursue this suggestion at some time in the future.

CORRIGENDA

Demonstration, "A Frank and Hertz critical potential experiment ", by
J. H. SANDERS (*Proc. Phys. Soc.*, **57**, 577 (1945)).

Page 578. In the wiring diagram of the oscillator (figure 2) the words "GRID" and
"FILAMENT" should be interchanged.

"The transient flow of heat through a two-layer wall", by MIRIAM V. GRIFFITH
and G. K. HORTON (*Proc. Phys. Soc.*, **58**, 481 (1946)).

Page 481, equation (1) : for d read ∂
Page 482, equations (4) and (5) : for d read ∂.
Page 482, last line :

$$\text{for} \quad \gamma = \frac{k_2 S_2 + k_1 S_1}{k_2 S_2 - k_1 S_1} \quad \text{read} \quad \gamma = \frac{\sqrt{k_2 S_2} + \sqrt{k_1 S_1}}{\sqrt{k_2 S_2} - \sqrt{k_1 S_1}}.$$

Page 483, equation (14) : transpose factors, to read

$$(2H/p)\sqrt{D_1 D_2/p} \{\lambda/(1 + \gamma e^{2x\sqrt{p/D_1}})\} \exp [a\sqrt{p}(D_1^{-\frac{1}{2}} + D_2^{-\frac{1}{2}})].$$

Page 483 equation (15) : insert a factor γ in the denominator of the term occupying the
second line.
Page 483 equation (16) : to read

$$\frac{2H\sqrt{D_1 D_2}\lambda}{\gamma p \sqrt{p}}\left(\frac{1}{1 + \gamma^{-1}e^{-2x\sqrt{p/D_1}}}\right) e^{\sqrt{p}}\left[x\left(\frac{1}{\sqrt{D_1}} - \frac{1}{\sqrt{D_2}}\right) - \frac{x}{\sqrt{D_1}}\right]$$

Page 483 : at end of the second of the four lines occupied by equation (17), insert $\Big)$.

OBITUARY NOTICES

MARIE PAUL AUGUSTE CHARLES FABRY

PROF. CHARLES FABRY was born at Marseilles on 11 June 1867. Though many of his working years were necessarily spent in Paris, he belonged to the South of France and, when opportunity allowed, returned there again and again. In a brilliant career at the Ecole Polytechnique he followed the example set by his grandfather, a pupil of Ampère. When later he became Membre de l'Institut, in the section of General Physics, he was following the lead given by two elder brothers, Eugène and Louis, who were already correspondents in Geometry and Astronomy, respectively.

Charles Fabry secured his *agrégation* in Physics in 1889 and his doctorate in Physical Sciences in 1892 with a thesis on the "Théorie de la visibilité et de l'orientation des franges d'interférence". Two years later, after a spell of teaching physics in several Lycées, Fabry was called back to Marseilles as Chief Lecturer in Physics under Professor Macé de Lépinay. Here with Perot, already in charge of the teaching of industrial physics, he formed a very fruitful partnership in research and a number of important papers in interferometry and spectroscopy were published : a series on the application of fringes produced by silvered plates led to the well known Fabry-Perot interferometer described in the *Annales de Chimie et de Physique* for 1901. Perot left Marseilles for a post in Paris and was succeeded by Buisson ; then came a succession of important papers on absolute wave-length in the spectrum of iron and cadmium and on the establishment of a new system of standard wave-lengths in the laboratory and the Sun ; this met a need long felt not only in the laboratory but also in astrophysics. Although Perot, Benoît and Buisson collaborated in these papers, the sure hand of Fabry was present throughout.

In 1904 Fabry succeeded Macé de Lépinay in the Chair of Physics at Marseilles, and while continuing his teaching in optics made a great name for himself locally as a teacher of industrial electricity ; his lucidity, simplicity and humour drew overflowing crowds of students, engineers and workpeople to his Wednesday evening lectures. His own work in the electrical field lay in absolute measures, as in optics. But in the latter field along with absolute values Fabry also studied with the aid of the interferometer very small displacements of spectral lines, for instance between the spectra of the Sun at the centre and at the limb ; also he applied his objective interferometer to differences in the line-of-sight velocities in different parts of the Orion nebula and to a study of the temperatures of the luminous gases.

In 1914 the first World War broke into his life and he left Marseilles to work with a radiographic unit in the field; later he became, under Paul Painlevé, technical adviser to the Artillery in the National Office of Research and Inventions : this led to his being sent as head of a scientific mission to the United States, where he was already so well known and so popular that he was welcomed everywhere and given every facility in his task. His position as one of the world's leaders in optics was so clear that he was the inevitable choice as the first director of " L'Institute d'Optique " when it was founded in Paris in 1920. At the same time he became Professor of Physics at the Sorbonne with Prof. Cotton as his colleague. His work for the optical industry of France, both in the class-room and in the optical workshop and factory, is reflected to-day not least in the rise of a brilliant school of French astrophysicists, his devoted pupils, and in the development of new and effective optical instruments.

The Thomas Young Oration given to our Society in 1935 on " Vision in Optical Instruments " is typical of this side of Fabry's scientific life : it shows his interest in practical considerations as modifying the results of classical theory and his recognition of the part played by the eye when used as an element of the optical train.

Fabry's other address to our Society, the Guthrie Lecture, delivered in 1925 on " The absorption of Radiation in the Upper Atmosphere ", represents another of the main branches of his scientific activity—spectrophotometry and photometry. The lecture is mainly devoted to the evidence for the absorption of the solar radiation by ozone at high levels

of the atmosphere : his study of this problem led to his presidency of the *Société Météoro-logique de France*. Other spectrophotometric studies include an examination of the distribution of the solar spectrum in the range λ 3150 A. – λ 2900 A.

From boyhood, when he used to observe eclipses of Jupiter's satellites, Fabry was always interested in astronomy. His applications of interferometry to astrophysics have already been mentioned, but he also worked on direct astronomical photometry, tackling such problems as the brightness of the night sky and of the corona and the absolute magnitude of the Sun in the stellar magnitude scale. His report on " La photométrie astronomique et celle des physiciens " is a most illuminating exposition of the principles underlying the work of the astronomer and the physicist and of the precautions which they severally and jointly must adopt. Fabry also gave the George Darwin lecture to the Royal Astronomical Society in 1938, and his subject, " Interstellar Space ", links up with his general interest in radiations of all types.

Recognition of Fabry's qualities as a scientific leader came freely from his colleagues all the world over. He was Vice-President of the International Astronomical Union from 1932 to 1935, and again from 1938 to 1945, serving on several of its scientific commissions and being President for many years of its Commission on Instruments. He was President of the International Council of Scientific Unions from 1937 until he resigned for reasons of ill-health shortly before his death. He was a very welcome attendant at many international gatherings where his lucidity, good sense and quick humour were invaluable and widely appreciated. A great lover of his country, he was sadly aware in 1939 of the awful threat hanging over it and of his country's state of unpreparedness to meet the coming danger. The years that followed must have been sad ones for him, but happily he survived to see France liberated and taking the first steps towards recovery.

He received in 1918 the Rumford medal of the Royal Society, of which he became a Foreign member in 1931. In 1921 he received the Franklin Medal of the Franklin Institute and was elected as a member. He was made an Associate of the Royal Astronomical Society in 1915, received the Draper medal of the National Academy of Sciences in 1919, and was elected a member of the Royal Institution in 1920. He became an Honorary member of the Physical Society in 1926.

He died in Paris on 11 December 1945, after an illness of several months.

F. J. M. STRATTON.

EVAN JOHN WILLIAMS, F.R.S.

E. J. WILLIAMS was born in the heart of rural Wales in 1903, the son of a master stone-mason. Educated at Llanwenog National School and then at Llandyssul County School, he took a scholarship which carried him to the University College, Swansea. Later he worked at Manchester under W. L. Bragg and then at Cambridge under Rutherford. After a year at Copenhagen with Bohr, he returned, first to Manchester and then to Chadwick at Liverpool. He was elected to the chair of Physics at Aberystwyth in 1936, and was elected into the Royal Society three years later.

He carried out both experimental and theoretical work on atomic and electronic collision processes. With the outbreak of war, his attention was turned to the immediate applied problems of the U-boat warfare, and, to quote Prof. Blackett's notice in *Nature* (156, 655). " It was largely due to Williams' keen analysis and powerful advocacy that an augmented and improved ' Bay offensive ' was staged in early 1943. The results, in terms of U-boats sunk, were not only in startling agreement with his predictions, but also were one of the important factors that led to the defeat of the U-boat campaign in 1943, and so made possible the serious planning of the invasion of Europe ".

(Based on material kindly made available by Prof. P. M. S. BLACKETT.)

IN the appreciation of E. J. Williams' life and work which appeared in *Nature* in December last, Professor Blackett has spoken of his brilliant contributions to the theory of collision problems ; he was also able, from first-hand knowledge, to do justice to the services which Williams rendered during the war in combating the U-boats. I did not know him so well in later years, but I can supplement Professor Blackett's account by my knowledge of his earlier career. He came to Manchester in the 'twenties as a young research student from University College, Swansea, with a warm recommendation from

Professor Evans. Our first impression was one of his blazing vitality and energy. I well remember the day of his arrival, because it happened to be the day of the annual rag between Owens College and the College of Technology, when a storming of each other's premises was a recognized procedure. Williams turned up in the laboratory next morning with his head swathed in bandages ; he had lost no time in entering into local politics. He was unconventional in the extreme. One usually knew where Williams was to be found in the laboratory from his habit of singing oratorios at the top of his voice when he felt like it, from sheer exuberance. To be driven by him in a car was a terrifying experience. He knew to a hair how close he could shave a tram, and would cut in between two advancing towards each other with a remark that it was really safer to do it that way because his brakes happened not to be working. He entered into everything he did with the same vast energy and enthusiasm. He and I collaborated in some work on the order-disorder phenomenon. It started at an afternoon colloquium, when I put forward some very tentative ideas in explanation of results on alloys which Bradley had been describing. Next morning Williams came to me with pages of calculations, embodying a well-developed mathematical treatment of the whole effect. It was so with all he did : he had an immediate and instinctive grasp of the essentials of a problem. Looking back now, I realize that we did not appreciate his full stature in those early days. He had a native genius for mathematical physics, brilliant but untutored, which needed for its full development contacts which we could not provide. He was at the same time no mean experimentalist, but his theoretical intuition made him impatient with the slow progress of experimental work ; his subsequent period at Cambridge was not as fruitful as it might have been because the division of energy between his experimental and theoretical interests hindered his deploying his full powers. Great though his contributions to science have been, one cannot but believe that had he lived he would have risen to still greater heights, and gathered a notable school of researchers round him in his own department. His death, as Professor Blackett says, leaves a sad gap in the ranks of our theoretical physicists. It is tragic that a man of such exuberant vitality and brilliant originality should have died so young. W. L. BRAGG.

JOHN LOGIE BAIRD

J. L. BAIRD, a Fellow of the Physical Society since 1927, died at his home at Bexhill on 14 June 1946 after a few months' illness, previous to which he had been actively engaged in experimental research in television in the laboratories of his own company.

Baird was the son of a Scottish minister and received his scientific education at the Royal Technical College, Glasgow. He first started on a business career, but later abandoned this to devote himself to an experimental study of the problems of television, which occupied him for the remainder of his life. Working under conditions of great difficulty due to limited means and apparatus, Baird correctly appreciated the basic problem of television, which consists in the provision of means for scanning an image by subdividing it into tiny elements, and transforming the resulting light variations into electrical impulses for transmission by line or radio to the receiver, where the impulses are converted back into light for the reconstruction of the picture. Using a mechanical scanning device and very intense illumination of the picture to be transmitted, Baird gave, in January 1926, what is claimed to be the first practical demonstration of television. This was followed in May 1927 by a demonstration of reception in Glasgow of picture images transmitted from London, and in February 1928 by the successful transmission of television across the Atlantic.

After these tests Baird Television Ltd. was formed to develop the technique for practical application, and the first step towards the inauguration of a public television service in Great Britain was taken in 1929, when the B.B.C. decided to give the above company facilities for experimental transmissions through the medium-wave London broadcasting station. Later the system was improved and adapted to operate on ultra-short waves, and towards the end of 1936 a public service was started from Alexandra Palace and comparison tests were conducted with the Baird and Marconi-E.M.I. systems, the latter being the one ultimately selected and now in use at the recently re-opened London Television Station.

During all these years, and until a few months before he died, Baird continued to work steadily towards the improvement of the scope and possibilities of television and its presentation. He had early taken advantage of the possibilities of the cathode-ray tube for reception,

and also demonstrated the optical projection of the television pictures on to a cinema screen The possibilities of introducing colour and stereoscopic effects also aroused his interest, and in 1944 he gave a demonstration of his recent achievements in the reception of television in colour by a method which avoided the need for revolving disks and lenses.

To some extent Baird's early work suffered some discredit and lack of recognition as a result of the exaggerated publicity and premature claims given to it by his supporters, but Baird himself was a shy and modest man, and there is no doubt that his pioneer achievements were a great incentive to other workers, and contributed materially towards the success attained by radio engineers and physicists, placing Britain in the forefront in this fascinating application of electromagnetic waves.   R. L. SMITH-ROSE.

GERALD SYDNEY FAWCETT

G. S. FAWCETT won for himself a unique position in colour physics. For nineteen years, until his death after a short illness on 8 November 1945, he was managing director of The Tintometer Ltd., the firm founded by his grandfather, J. W. Lovibond. In this position it was his constant desire to see his grandfather's invention used and developed to the benefit of his fellow men. Recent developments which enabled the Tintometer to be used in medical science were a source of particular satisfaction to him.

Born in Salisbury in 1896, he was educated at the Modern School, Salisbury, and at Christ's Hospital. Between 1915 and 1918 he served in the R.A.M.C. and in the Labour Corps. He then trained and practised for a short time as a quantity surveyor. In 1925 he married Marion, daughter of Mr. and Mrs. Oscar Wheater. She and three children, Paul, Janet and Joe, survive him.

As a boy he took a great interest in his grandfather's work, and he gained his wide knowledge of colour physics through individual study and practical experience rather than from systematic instruction. He became a Fellow of the Optical Society in 1926 and of the Physical Society in 1932. He joined the Colour Group when it was formed in 1940, and was a member of its committee at the time of his death. He was joint author with Prof. R. H. Stoughton of *Chemical Testing of Plant Nutrient Solutions*.

To know Gerald Fawcett was to know a friend, and what a good friend he was ! It was a cruel fate that took him from us so soon, but our memories of him remain an inspiration.   R. K. S.

DANIEL EVAN JONES

By the death of D. E. JONES the Physical Society has lost yet another of the fast-dwindling band of Fellows who joined its ranks in the first fifteen years of its existence. He was elected to the Society in 1886.

After a period of study at the University College of Wales, Aberystwyth, and at Owens College, Manchester, under Sir Henry Roscoe and Balfour Stewart, he took his B.Sc. degree in 1885 at London University with first-class honours in chemistry. For some years he held the chair of physics at University College, Aberystwyth, but, feeling that his gifts lay rather in the direction of administration, he left the teaching profession for the posts of Director of Technical Education to the Staffordshire County Council and Regional Inspector under the Department of Science and Art. He was the author of an elementary textbook on Heat, Light and Sound, and of a set of examples in physics ; these books had a considerable vogue some forty years since.

But Jones's major contribution to physics, and one by which he will be most gratefully remembered, consists in his lucid and scholarly translation of the works of Heinrich Hertz. In 1893 he published, under the title of *Electric Waves*, translations of Hertz's epoch-making papers. The volume was introduced by a very characteristic preface, which is still of historical interest, from the pen of Kelvin. In 1896 appeared a translation, made in conjunction with G. A. Schott, of Hertz's *Miscellaneous Papers*, and in 1899, with the assistance of J. T. Walley, a translation of Hertz's *Principles of Mechanics*. The two last-named volumes contained most valuable biographical and critical introductions by Lenard and by Helmholtz, and the three volumes were of great value in making the work of Hertz accessible to British physicists. They are the translator's most enduring

... of the Society, and is
A. F.

REVIEWS OF BOOKS

Dictionary of Science and Technology in English-French-German-Spanish, by MAXIM NEWMARK. Pp. viii + 386. (London: Sir Isaac Pitman and Sons, Ltd., 1945.) 30*s.*

The author is in the Modern Language Department of the Brooklyn Technical High School, and has experience both of teaching technical students and of translation. He began his book as notes from experience, and then systematically extended it by compiling word lists from the indexes to textbooks and by examination of other technical dictionaries. It is arranged with the English words in alphabetical order, followed by the three equivalents. Each such English word has a number, and there are indexes of French, German and Spanish words, giving reference to the initial letter and number of the corresponding English word. Each continental word has its gender set out in the main table, a feature much to be commended. It is a little difficult, however, to see precisely the class of user to whom the dictionary is most intended to appeal. A translator from English to one of the other languages finds the words he requires at once, with their genders, but he hardly needs the non-technical words which are so copiously supplied (across, to act, broom, brush, electric, index, nature, under, zig-zag and zinc). The translator, and indeed the general reader who is competent to undertake the perusal of matter in French, German or Spanish, could equally well do without most of these words, or at any rate could take them from an ordinary dictionary in the rare cases where a peculiarity of use of prepositions is in question, but he can hardly need the words which are alike in all languages. Thus we have on a single page

Macadam	macadam	Makadam	macadám
Machine	machine	Maschine	máquina
magnesia	magnésie	Magnesia	magnesia
magnesite	magnésite	Magnesit	magnesita
magnesium	magnésium	Magnesium	magnesio

and on another

cupreous	cuivreux	Kupferhaltig	cúprico
cyanamide	cyanamide	Zyanamid	cianamido
cyanide	cyanure	Zyanid	cianuro
cycle	cycle	Zyklus	ciclo
cycloid	cycloïde	Zykloide	cicloide,

though it is true that on this page we also find the entry " cyclotron ", " same in all languages ".

As to the actual coverage, there are 10,000 entries and the book has a vast number of workshop terms, a great deal of physics and chemistry, and practically no biology. The word " bee " does not appear, nor does " owl ", but " wheat " and " oats ", " birch " and " beech " do, though " barley ", " corn " and " maize " do not. Notable omissions are " bulldozer " and " fission ". There are a few surprises, most of which reflect rather the difference between English and American than any error on the part of the author. Thus " Adobe " translates into French " Argile " and German " Tonerde ", and " clay " into " argile " and " Lehm ", whereas " mud " is " Bourbe " and " Schlamm " ; under " navvy " we read " same as *steam shovel* ". One meaning given for *airline* is " à vol d'oiseau ", which in this country could only be expressed as " bee line " or " as the crow flies ". " Police " appears, but not " Policeman ", so that " Agent " in the French index misses one important meaning, and the familiar " Schupo " does not appear at all. Finally, without any suggestion that there is an omission or error, we may sympathize with the French on learning from this dictionary that every time they wish to speak of a radio beam they have to write or say " rayon radiogoniométrique ".

J. H. A.

Relaxation Methods in Theoretical Physics, by R. V. SOUTHWELL.   Pp. vi + 248.
    (Oxford: The Clarendon Press, 1946.)   20s.

The value of the relaxation methods introduced some years ago by Professor Southwell
is by now well known, and it is generally realized that they are not confined to problems
in statics.   In the present volume, it is shown that these methods may be extended to deal
with the solution of boundary-value problems associated with partial differential equations
in two independent variables.   For this purpose, the equation is replaced by its equivalent
in finite differences, and a network covering the region of space concerned is set up, on
which the values of the dependent variable can be computed by a method in which the
boundary conditions are complied with, and the equation is satisfied by adjustments similar
to those used in the more familiar relaxation method.

The basis and the methods used are both explained clearly, and many solved examples
are given, so that the reader, even if disinclined to work out the whole problem for himself,
may verify quite quickly that the solution given does in fact satisfy the required differential
equation.   As was remarked in the review of the earlier book, *Relaxation Methods in
Engineering Science* (and more neatly by a reviewer quoted with approval by the author),
it is essential for a reader who really wishes to grasp the subject to perform a few com-
putations himself.   Even without this, however, it is possible to appreciate the value of the
method to some extent, and to admit that when it can solve problems of which the solution
has hitherto been sought in vain, it must be worthy of attention.   Several such problems
are dealt with, including the flow of liquid under gravity to form a waterfall.

With such a record of success, and with an exposition which would in any case pre-
dispose a reader in its favour, it seems a pity that the author has thought it necessary to
write about the method in terms which would suggest that there is a conspiracy against it,
and that it is a rival (instead of a supplement) to orthodox methods, the practitioners of
which are attempting to suppress it.   Thus we are told that the aims of orthodox analysis
of a physical problem are to obtain general and exact solutions, and it is implied that
relaxational methods have different aims.   Now, in so far as the new methods solve only
one case of a given problem (for example, one geometrical parameter may have to be fixed),
it is surely indisputable that a general solution would be preferable.   But as regards
accuracy, it seems to the present reviewer that the relaxationists and the followers of
orthodox analysis both have the same aim—to solve the problem within pre-specified
limits.   The former do it numerically, the latter may give, for example, a convergent
expansion, of which as many terms are worked out as may be needed, or they may give an
asymptotic expansion, in which case care is taken to see that the possible error is less than
the pre-specified figure for those values of the variables which are in question.   Again,
a good deal is said about the meaning attached by the theoretical and the practical man to
the word " function "; but in fact nothing which is said here would surprise a reader of
orthodox treatises, particularly of finite differences or calculus of observations, save in the
section on basic theory of conformal transformation, where " function " suddenly takes
on the meaning assigned by the rigourist to " analytical function ", and does not have the
meaning so carefully set out for it by the author.

This " Heaviside complex " apart, the book is a model of good exposition, on a subject
of real importance.   It is beautifully produced, and the publishers deserve commendation ;
their task was not easy, for in order to give the solutions of problems in a legible manner,
it has been necessary to include about a dozen folding plates, as well as to set up a great
deal of mathematical and tabular matter.   J. H. A.

Electric Discharge Lighting, by F. G. SPREADBURY.   Pp. viii + 136, 122 figures.
    (London: Sir Isaac Pitman and Sons, Ltd., 1946.)   15s.

Rapid progress has been made in the development of electric discharge lamps during
the past 10 or 15 years, and a great variety of types, from simple negative glow lamps dissi-
pating only a few watts to high pressure mercury vapour lamps dissipating several kilo-
watts, have been developed.   In some of these lamps the discharge itself is the light source,
while in others the light is obtained from a layer of luminescent material excited by the
ultra-violet radiation from the discharge which contributes only a negligible amount of

direct light. The importance of these lamps lies mainly in the fact that the electric discharge permits of much greater possibilities as regards luminous efficiency and control of colour than does the tungsten filament lamp.

In the book under review the author has, somewhat ambitiously, attempted to deal with all types of discharge lamps in the space of some 130 pages. Following an introductory chapter describing certain theoretical aspects of radiation and electric discharges, there are three chapters, of some 60 pages in all, dealing with the lamps themselves and including a section on luminescence. The lamps described are those developed by the leading lamp makers in this country and no attempt has been made to deal with American electric discharge lamp practice. Although the book claims to deal with all types of discharge lamps in current use there is no mention of the important and interesting compact source mercury vapour lamps, which were largely developed in this country. Two chapters are devoted to circuit and gear matters, while there is a somewhat irrelevant concluding chapter on *Technical Applications of Discharge Lamps* ; it comes as something of a surprise to find the use of the gas discharge triode described in this chapter.

It is not clear for what type of reader the book is particularly intended : workers in the field will hardly find anything new or stimulating in what is, after all, little more than an assembly of well-known information. The lighting engineer is likely to be disappointed in the book for, despite its title, lighting problems are simply not discussed. It appears probable, however, that the book is not addressed primarily to either of these categories, but rather to the general reader who wants less specialized accounts of particular fields of activity, and to this class of reader it is recommended.

The book is excellently produced and printed, and the author has managed to assemble a considerable amount of material within the compass of a very compact volume. The diagrams and illustrations are excellent and the reproduction of oscillograms of voltage and current wave forms is particularly pleasing. A useful feature of the book is the collection of tables and curves giving performance and dimensional characteristics of British discharge lamps. It should be realized, however, that such a collection of data is liable to get out of date quickly—indeed some of the information given in the tables is already out of date—and it is always advisable to consult the lamp manufacturers for present-day information relating to their products.

In writing a book on practically any technical subject, the problems of acknowledging the contributions of other workers in the field naturally arises. The author of the present book solves the difficulty by avoiding acknowledgments. Without presuming to criticize this procedure, it may perhaps be suggested that the value of the work would have been greatly enhanced by the inclusion of a fairly full bibliography, which, in addition to assisting the reader who wants to probe deeper into the subject, would serve to indicate the source of much of the work described in the book. No doubt this is a matter which can be considered in connection with subsequent editions. H. G. JENKINS.

Mathematical Tables. Part-volume A, Legendre Polynomials, by the Committee for the Calculation of Mathematical Tables. Pp. 42. (Published for the British Association at the University Press, Cambridge, 1946.) 8*s.* 6*d.*

The volumes issued by the British Association Tables Committee ushered in the new era of plentiful, and reliable, tables, and these volumes are now well known to all who work in applied mathematics. In the past, the volumes have been of comparatively large size, and usually devoted to one type of table. The need has now been felt for the publication of various smaller tables, and to avoid waiting until enough disconnected material has been accumulated, the Association has adopted the device of issuing part-volumes.

The issue now under review contains the Legendre polynomials $P_n(x)$ regular at the origin. When the work was started, there were three tables in existence—the famous one of Perry up to $n=7$, and two up to $n=8$, all for $x=0$ to 1. The present table carries n up to 12, and in this range of x all are correct to at least 7 figures. There were at the inception of the project no tables beyond $x=1$, though in the interval several have appeared (Tallqvist, 1937 ; Prevost, 1933 ; Dwight, 1945). In the present volume, x is taken to 6 for all the values of n given, and to 11 for the first six values of n.

The tables are thus a valuable complement to those previously available. They are well printed and carry a clear and useful introduction by L. J. Comrie, whose name on a book of tables is almost a guarantee that there are no errors. J. H. A.

Mathematical Tables. Part-volume B, the Airy Integral, by J. C. P. MILLER. Pp. 56. (Published for the British Association at the University Press, Cambridge, 1946.) 10s.

The complete solution of the equation $y''=xy$ is $y=A\,\mathrm{Ai}(x)+B\,\mathrm{Bi}(x)$, or in an alternative form, $y=CF(x)\sin[\chi(x)+\epsilon]$. With a linear change of both variables, the function $\mathrm{Ai}(x)$ gives the integral $\int_0^\infty \cos\frac{1}{3}\pi(w^3-mw)\,dw$, whilst both Ai and Bi are closely related to modified Bessel Functions of order $\pm1/3$. In the present part-volume, Ai and its first derivative are given for $x=-20$ to $+2$ at intervals of $0\cdot01$, and Bi from -10 to $+2\cdot5$ at intervals of $0\cdot1$. The logarithms and logarithmic derivatives of both are also given, the former from 0 to 25 at intervals $0\cdot1$ and thence to 75 at unit intervals, the latter from 0 to 10 at intervals of $0\cdot1$. The first 50 zeros and turning values of Ai and the first 20 of Bi are also given, together with tables of F and χ and the corresponding functions for the derivatives.

In addition, as we have come to expect from the B.A. Tables, there is a useful introductory treatise on the mathematics of these functions, with an elaborate account of the computational methods used. J. H. A.

Electrons in Action, by J. G. DAUNT. Pp. 151. (Sigma Books, 1946.) 6s.

In the preface of this book the author states that his purpose is to give the non-scientific reader an understandable account of electricity in its various aspects without recourse to mathematical symbols or formulae. He succeeds in the avoidance of technical algebra, except for the mention of Ohm's Law in the appendix, and here the use of the symbol C for current is perhaps a little unfortunate as being out of accord with modern practice. In making his subject intelligible to the ordinary reader Dr. Daunt has certainly spared no effort to find apt analogies with which to illustrate his points ; his style throughout is interesting, though as a whole it lacks the fascination of the Mr. Tompkins books.

The field of electro-technics covered by this book is very wide, the author commencing with a simple exposition of the electron theory of matter and, with these fundamental ideas established, the reader is introduced in turn to the phenomena of conduction of electricity, electro-magnetism and electric heating. Later chapters deal with thermionic emission, x rays, photo-electricity, radar, cosmic rays, etc., and finally the author introduces the idea of the dualism of matter and Heisenberg's *principle of uncertainty*.

The work contains a number of photographic plates appropriately chosen, together with a large number of line drawings which, though often quite original in conception, are in some cases marred by poor execution. This book should prove a most useful introduction to the subject for non-scientific school-pupils and adults alike, and for those readers who are sufficiently stimulated to read further, Dr. Daunt has thoughtfully inserted, in the appendix, a list of books of graded difficulty. R. W. B. S.

Alternating Current Measurements at Audio and Radio Frequencies, by DAVID OWEN. Pp. 120+vii, with 80 diagrams, Second Edition, revised. (London: Methuen and Co., Ltd., 1926.)

This book is a second and revised edition of a popular monograph of the Methuen series. In it the general features of the original have been retained, the opening chapter containing the requisite A.C. theory for the clear understanding of the various bridges, potentiometers and resonance circuits described in succeeding chapters. The experimental procedure and the calculation of results from typical observations are given with each method described, the clarity of the explanation in the text being enhanced by

means of line diagrams. The chapter dealing with the conditions of accuracy in bridge measurements at low frequencies is most instructive and useful, and in fact the only note of criticism is in regard to the section on radio-frequency measurements. The author has neglected here an opportunity to bring the text right up to date, for contrary to his remarks, bridge methods at these high frequencies are now actually in commercial use.

R. W. B. S.

Optique Instrumentale, by G. A. BOUTRY. Pp. x + 539. (Paris: Masson et Cie, 1946.)

Professor Boutry's charming book represents the course of instruction on the theory of optical instruments given by him to students at the Institut d'Optique. It is not concerned with the problems of optical design, but first examines the fundamental theorems on which all sound optical designing is based, and passes on to consider the construction and use of representative instruments of modern types.

As the author explains in an interesting preface, his aim is to present a clear picture of the subject—still living and growing—rather than to adhere to any pedantic limitations, although it must be said that little exception can be taken to the general foundation. The mathematics takes, however, a generally subordinate position, the main effort being towards clarity of exposition in which the use of frequent clear diagrams will appeal to most students. The graphical presentation of the changes in spherical aberration of a refracting surface according to conjugate distance can be mentioned as a praiseworthy sample of the use of the diagrams to convey in one glance information which can only come from hard work with a formula.

There is a very useful and up-to-date chapter on the Eye and Vision in relation to the use of optical instruments. The well-known researches of Arnulf and others at the Institut d'Optique are freely drawn upon for material here.

If there can be any general criticism (apart from detail) one might feel this general picture of the edifice of optical knowledge is almost too well-drawn. There are only " one or two " references to original papers, and even references to proper names are relatively few, and mostly restricted to classical authors.

In the discussion of the theory of the microscope, for example, the reader would get no impression that this subject has been a battle-ground of strong controversy, and the treatment in the book still seems to lead to erroneous ideas regarding the proper function of the condenser in microscopy ; it introduces once more (though not in the same words of course) the old error of critical illumination.

The time is passing, perhaps, when a student expected a textbook to be verbally inspired on every detail. While directness of statements is generally to be welcomed, it should not prevent an author from indicating the points where the discussion is still incomplete and further enquiry is necessary. A book may be like a perfectly painted picture by van Eyck ; one leaves it in admiration—but an unfinished Michael Angelo may provide more stimulation to the imagination. In spite of the author's professed aim in the preface, he has not always been successful in indicating the " lines for development ", though a great many comments throughout the book will undoubtedly prove stimulating and helpful to any optically-minded reader ; probably he feels that the lines for development " belong more to the province of the designer than to the general student of optical instruments, but even so, the attractive possibilities of systems with aspheric surfaces—including the Schmidt camera—might have been expected to receive a fuller discussion.

Professor Boutry has asked that a minor defect (on page 217) in the treatment of systems not possessing full axial symmetry should be pointed out : a statement appropriate to a single surface has been inadvertently generalized to cover a complete system. On the whole, however, the book seems very free from errors.

Its sign convention will not, perhaps, appeal to the majority of British readers ; it is not in accordance with the systems which have been most favoured in this country. There must be very few, however, who could peruse the book without adding most usefully to their optical knowledge.

L. C. MARTIN.

INDEX

INDEX TO REVIEWS OF BOOKS